Estimation and Inference in Large Heterogenous Panels with Cross Section Dependence

M. Hashem Pesaran
Trinity College, Cambridge

October 2002

Abstract

This paper presents a new approach to estimation and inference in panel data models with unobserved common factors possibly correlated with exogenously given individual-specific regressors and/or the observed common effects. The basic idea behind the proposed estimation procedure is to filter the individual-specific regressors by means of (weighted) cross-section aggregates such that asymptotically as the cross-section dimension (N) tends to infinity the differential effects of unobserved common factors are eliminated. The estimation procedure has the advantage that it can be computed by OLS applied to an auxiliary regression where the observed regressors are augmented by cross sectional averages of the dependent variable and the individual specific regressors. It is shown that the proposed correlated common effects (CCE) estimators for the individual-specific regressors (and its pooled counterpart) are asymptotically unbiased as $N \to \infty$, both when T (the time-series dimension) is fixed, and when N and T tend to infinity jointly. Further, the CCE estimators are asymptotically normal for T fixed as $N \to \infty$, and when $(N,T) \to \infty$, jointly provided $\sqrt{T}/N \to 0$ as $(N,T) \to \infty$. A generalization of these results to multi-factor structures is also provided. The estimation and inference in dynamic heterogenous panels with a residual factor structure will be addressed in a companion paper.

Keywords: Cross Section Dependence, Large Panels, Common Correlated Effects, Heterogeneity, Estimation and Inference.

JEL-Classification: C12, C13, C33.

1 Introduction

A number of different approaches have been advanced in the literature for the analysis of cross section dependence. In the case of spatial problems where a natural immutable distance measure is available the dependence is captured through "spatial lags" using techniques familiar from the time series literature. In economic applications spatial techniques are often adapted using alternative measures of "economic distance". See, for example, Lee and Pesaran (1993), Conely and Dupor (2001), Conley and Topa (2002) and Pesaran, Schuermann and Weiner (2001), as well as the literature on spatial econometrics recently surveyed by Anselin (2001). In the case of panel data models where the cross section dimension (N) is small (typically $N < 10$) and the time series dimension (T) is large the standard approach is to treat the equations from the different cross section units as a system of seemingly unrelated regression equations (SURE) and then estimate the system by the Generalized Least Squares (GLS) techniques. This approach allows for a general (time-invariant) correlation patterns across the errors in the different cross section equations.

There are also a number of contributions in the literature that allow for time-varying individual effects in the case of panels with homogenous slopes where T is fixed as $N \to \infty$. Holtz-Eakin et al. (1988) use a quasi-differencing procedure to eliminate the time-varying effects and then estimate the model by instrumental variables. This procedure eliminates the individual-specific effects but yields (dynamic) regression equations with time-varying coefficients that are generally difficult to estimate and is likely to work only when T is quite small. Ahn, Lee and Schmidt (2001), building on the earlier contributions of Kiefer (1980) and Lee (1991) propose a number of different generalized method of moments (GMM) estimators depending on whether first as well as second-order moment restrictions are utilized. In the case where idiosyncratic errors are homoskedastic and nonautocorrelated, they show that the GMM estimator that makes use of all the first- and second-order moment restrictions dominates the maximum likelihood estimator (which is also the generalized within estimator) originally proposed by Kiefer (1980). However, their analysis assumes that the regressors are identically and independently distributed across the individuals, which may not be valid in practice. In addition, none of these approaches are appropriate when both N and T are large and of the same order of magnitudes, as is often the case in cross-country (region) studies.

The application of unrestricted SURE-GLS approach to large N and T panels involves nuisance parameters that increase at a quadratic rate as the cross section dimension of the panel is allowed to rise. To deal with this problem Robertson and Symons (2000) propose restricting the covariance matrix of the errors using a common factor specification with a fixed number of unobserved factors, and then estimating the model using maximum likelihood (ML) techniques assuming homogeneous slopes. Robertson and Symons also assume that the common factors are distributed independently of the observed regressors that are included in the model. In this case and under certain restrictions on the number of fitted factors Robertson and Symons show that the likelihood function is bounded and the maximum likelihood estimates exist even in the rank deficient case where $N > T$. However, they are unable to establish the asymptotic properties of their estimators in the case of large N and primarily rely on results from Monte Carlo experiments in the context of a simple panel data model with homogeneous coefficients. The SURE-GLS procedure is also utilized recently by Phillips and Sul (2002) for estimation of autoregressive models with heterogeneous slopes (but without exogenous regressors) using a single factor structure for the residuals. Once again no large N asymptotic results are provided. In another related paper Coakley, Fuertes and Smith (2002) propose a principal components approach that is arguably simpler to implement than Robertson and Symons's full ML procedure.[1] These authors also claim that their procedure is valid even if the unobserved common factors and the observed individual effects are correlated, possibly due to omitted global variables or common shocks that are correlated with the included regressors.

In this paper we first establish that in general the estimation procedure proposed by Coakley, Fuerts and Smith (CFS) will not be consistent if the unobserved factors and the included regressors are correlated. We also show that the satisfactory simulation results reported in the paper is due to the paper's special Monte Carlo design where the cross-section average of the included regressor and the unobserved common effect become perfectly correlated as $N \to \infty$. We shall then propose a new approach that yields consistent and asymptotically normal parameter estimates even in the presence of correlated unobserved common effects both when T is fixed and $N \to \infty$, and as $(N, T) \to \infty$, jointly.

Initially, we consider a one-factor residual model and distinguish between individual-specific regressors, observed and unobserved common effects. We permit the common effects to have differential impacts on individual units, while at the same time allow them to exhibit an arbitrary degree of correlation amongst themselves and with the individual-specific regressors. We consider the problem of estimating individual slope coefficients as well as a pooled estimator when the coefficients of individual-specific regressors are homogeneous. We allow for error variance heterogeneity and do not assume that the individual-specific regressors are identically and/or independently distributed over the cross-section units, that are particularly relevant to the analysis of cross-country panels. However, in this paper we assume the regressors to be stationary and exogenous. Due to space limitations, the problem of estimation and inference in dynamic panels subject to cross section dependence will be addressed in a companion paper. See Pesaran (2002).

The basic idea behind the proposed estimation procedure is to filter the individual specific regressors by means of (weighted) cross section aggregates such that asymptotically (as $N \to \infty$) the differential effects of unobserved common factors are eliminated. This is in contrast with the various approaches adopted in the literature that focus on estimation of factor loadings as an

[1] Similar issues are also discussed in the analysis of (dynamic) factor models by Forni and Lippi (1997), Forni and Reichlin (1998), Stock and Watson (1998), and Bai and Ng (2002), among others.

input into the GLS algorithm. The estimation approach has the added advantage that it can be computed by ordinary least squares applied to an auxiliary regression where the observed regressors are augmented by cross sectional averages of the dependent variable and the individual specific regressors. We refer to this estimator as the common correlated effects (CCE) estimator. We also propose a pooled CCE estimator which is asymptotically efficient when the slope coefficients of the individual-specific regressors are homogenous, although the common effects (whether observed or not) are still allowed to have differential impacts over the cross-section units. The pooled CCE estimator can also be viewed as a generalized fixed effects estimator where the individual-specific dummies are augmented with individual-specific cross-section aggregates.

We show that the CCE estimator of the coefficients of the individual-specific regressors (and its pooled counterpart) are asymptotically unbiased as $N \to \infty$, both when T is fixed, and when N and T tend to infinity jointly. Further we show that the proposed estimators are asymptotically normal for T fixed as $N \to \infty$, and when $(N, T) \to \infty$, jointly provided $\sqrt{T}/N \to 0$ as $(N, T) \to \infty$. A generalization of these results to multi-factor structures is also provided. In the case where T is fixed our proposed estimator while directly comparable to the GMM estimator proposed by Ahn et al. (2001), is applicable more generally. In particular, it does not require the individual-specific regressors to be cross sectionally independent.

Finally, it is worth emphasizing that the CCE estimator is applicable even if one is interested in individual-specific coefficients. In such a case not allowing for cross dependence would not pose inconsistency problems (only inefficiencies) if the source of cross dependence (the common effects) are not correlated with the observed included regressors. Seen from this perspective the paper addresses two problems: cross dependence and correlated omitted effects. It is of interest that it is possible to find a solution to the second problem by viewing individual regressions as parts of a fully integrated panel data model. In effect the approach proposed in the paper provides another generalization of Zellner's (1962) idea to the case of seemingly unrelated regressions where the cross-sectionally dependent errors are also correlated with the included regressors. The differences in the two approaches stem from what is assumed about the size of N. In Zellner's framework since N is fixed (and relatively small) cross-dependence can be modelled freely, but it is assumed that the unobserved errors and the included regressors are uncorrelated. In the case where N is large cross-dependence can only be modelled subject to restrictions, but the assumption of zero correlation between the errors and the included regressors can be relaxed.

The plan of the paper is as follows: Section 2 sets out the one-factor residual model and the assumptions in detail. Section 3 shows the general inconsistency of the principal components estimator proposed by Coakley, Fuertes and Smith (2002). Section 4 first motivates the idea of approximating the unobserved common factor by linear combination of the cross-sectional aggregates of the dependent and the individual-specific regressors. The CCE estimators of the coefficients of the individual-specific regressors are then presented in sub-section 4.1, and their pooled counterpart in sub-section 4.2. The problem of how best to choose the weights used in the construction of the aggregates and in the formation of the pooled estimator are discussed in sub-section 4.3. Section 5 provides a generalization of the CCE estimators in the case of a multi-factor residual model. Section 6 concludes by identifying important areas for extensions and further developments.

Notations: K stands for a finite positive constant, $\|\mathbf{A}\| = [Tr(\mathbf{AA}')]^{1/2}$ is the Euclidean norm of the $m \times n$ matrix $\mathbf{A}$. $a_n = O(b_n)$ states the deterministic sequence $\{a_n\}$ is at most of order b_n, $\mathbf{x}_n = O_p(\mathbf{y}_n)$ states the vector of random variables, $\mathbf{x}_n$, is at most of order $\mathbf{y}_n$ in probability, and $\mathbf{x}_n = o_p(\mathbf{y}_n)$ is of smaller order in probability than $\mathbf{y}_n$, $\overset{q.m.}{\to}$ denotes convergence in quadratic mean (or mean square error), $\overset{p}{\to}$ convergence in probability, $\overset{d}{\to}$ convergence in distribution, and $\overset{d}{\sim}$ asymptotic equivalence of probability distributions. All asymptotics are carried out under $N \to \infty$, either with a fixed T, or *jointly* with $T \to \infty$. Joint convergence of N and T will be denoted by $(N, T) \overset{j}{\to} \infty$. Restrictions (if any) on the relative rates of convergence of N and T will be specified separately.

2 A One-Factor Residual Model

Let y_{it} be the observation on the i^{th} cross-section unit at time t for $i = 1, 2, ..., N$; $t = 1, 2, ..., T$, and suppose that it is generated according to the following linear heterogeneous panel data model

$$y_{it} = \mathbf{z}_t' \boldsymbol{\alpha}_i + \mathbf{x}_{it}' \boldsymbol{\beta}_i + u_{it}, \tag{2.1}$$

where $\mathbf{z}_t$ is a $k_z \times 1$ vector of observed common effects, $\mathbf{x}_{it}$ is a $k_x \times 1$ vector of observed individual-specific regressors on the i^{th} cross section unit at time t, and the errors have the one-factor structure

$$u_{it} = \gamma_i f_t + \varepsilon_{it}, \tag{2.2}$$

in which f_t is the unobserved common effect, and ε_{it} are the individual-specific (idiosyncratic) errors. The primary parameters of interest are the individual specific slope coefficients, $\boldsymbol{\beta}_i$, $i = 1, 2, ..., N$. The common factor loadings, $\boldsymbol{\alpha}_i$ and γ_i, will be treated as nuisance parameters, although we shall also consider their identification and estimation once consistent estimators of $\boldsymbol{\beta}_i$ are obtained.

The following assumptions will be made throughout:

Assumption 1(a) (observed individual-specific regressors): For each i the regressors, $\mathbf{x}_{it}$, are covariance stationary with absolutely summable autocovariances, zero means and finite fourth-order moments and are distributed independently of the individual-specific errors, $\varepsilon_{it'}$, for all t and t'. In particular, for each t

$$E(\mathbf{x}_{it}' \mathbf{x}_{it})^s \leq K, \text{ for } s = 1, 2 \text{ and all } i. \tag{2.3}$$

Dependence between $\mathbf{x}_{it}$ and f_t is allowed so long as for each i as $T \to \infty$

$$\frac{1}{T}\sum_{t=1}^{T} f_t \mathbf{x}_{it} \xrightarrow{p} \sigma_{fi} < \infty,$$

and

$$\frac{1}{T}\sum_{t=1}^{T} f_t \bar{\mathbf{x}}_t \xrightarrow{p} \sigma_{f\bar{x}},$$

where $\sigma_{f\bar{x}} = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \sigma_{fi} < \infty$.

Assumption 1(b) (observed common effects): The observed common effects, $\mathbf{z}_t$, are covariance stationary with absolute summable autocovariances, distributed independently of the individual-specific errors, $\varepsilon_{it'}$, for all t and t'.

Assumption 1(c) (unobserved common effect): The single unobserved common effect, f_t, is covariance stationary with absolute summable autocovariances, mean zero and variance σ_f^2, and for each i , f_t and $\varepsilon_{it'}$ are independently distributed for all t and t'.

Assumption 2(a) (fixed factor loadings): The factor loadings, γ_i, are non-stochastic constants such that $|\gamma_i| \leq K$.

Assumption 2(b) (random factor loadings): The factor loadings, γ_i, follow the random coefficient model

$$\gamma_i = \gamma + \eta_i, \;\; \eta_i \backsim iid\ (0, \sigma_\eta^2), \text{ for } i = 1, 2, ..., N, \tag{2.4}$$

where $\gamma \neq 0$, $0 \leq \sigma_\eta^2 < K$, and $\eta_i' s$ are distributed independently of the regressors $(\mathbf{x}_{it}', z_t, \ f_t)$ for all i and t.

Assumption 3 (random coefficients): The slope coefficients of the individual-specific effects, $\boldsymbol{\beta}_i$, follow the random coefficient model

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{v}_i, \;\; \boldsymbol{v}_i \backsim iid\ (\mathbf{0}, \Omega), \text{ for } i = 1, 2, ..., N, \tag{2.5}$$

where Ω is a $k_x \times k_x$ non-negative definite matrix and the random deviations, $\boldsymbol{v}_i$, are distributed independently of ε_{jt}, the regressors $(\mathbf{x}_{jt}', \mathbf{z}_t', f_t)'$ for all i, j and t, and

$$E\left(\boldsymbol{v}_i'\boldsymbol{v}_i\right)^2 \leq K. \tag{2.6}$$

The coefficients of the observed common effects, $\boldsymbol{\alpha}_i$, are bounded (lie on a compact set).

Assumption 4 (individual-specific errors): The individual specific error, ε_{it}, is distributed independently across i and t with mean zero, variance, σ_i^2, and a finite fourth-order moment, $E(\varepsilon_{it}^4) \leq K$.

Assumption 5(a): (identification of unobserved common effects): There exists a set of fixed (aggregating) weights, $\{w_i, i = 1, 2, ..., N\}$, such that for all i and N (including as $N \to \infty$)

$$\text{(i): } w_i = O\left(\frac{1}{N}\right), \text{ (ii): } \sum_{i=1}^{N} |w_i| < K, \text{ and (iii): } \sum_{i=1}^{N} w_i\gamma_i \neq 0. \tag{2.7}$$

Assumption 5(b): (identification of $\boldsymbol{\beta}_i$): Let $\bar{\mathbf{x}}_{wt} = \sum_{j=1}^{N} w_j \mathbf{x}_{jt}$, with the weights $\{w_j\}$ satisfying the conditions of assumption 5(a), and consider the partition $\bar{\mathbf{x}}_{wt} = (\bar{\mathbf{x}}'_{1wt}, \bar{\mathbf{x}}'_{2wt})'$ associated with $\mathbf{x}_{it} = (\mathbf{x}'_{1it}, \mathbf{x}'_{2it})'$. Assume that $\mathbf{x}_{1it}$ exhibit a sufficient degree of cross section dependence such that for each t and as $N \to \infty$, $\bar{\mathbf{x}}_{1wt} \overset{q.m.}{\to} \boldsymbol{\mu}_{t,x_1} \neq \mathbf{0}$, a time-varying vector of random variables, and that $\mathbf{x}_{2it}$, are not sufficiently cross sectionally dependent and $\bar{\mathbf{x}}_{2wt} \overset{q.m.}{\to} \mathbf{0}$, as $N \to \infty$. Specifically

$$\bar{\mathbf{x}}_{1wt} - \boldsymbol{\mu}_{t,x_1} = O_p\left(\frac{1}{\sqrt{N}}\right), \tag{2.8}$$

and

$$\bar{\mathbf{x}}_{2wt} - \boldsymbol{\mu}_{t,x_2} = O_p\left(\frac{1}{N}\right). \tag{2.9}$$

Let $\mathbf{g}_{w,it} = (\mathbf{x}'_{it}, \mathbf{z}'_t, \bar{\mathbf{x}}'_{1wt}, f_t)' = (\mathbf{x}'_{it}, \bar{\mathbf{g}}'_{wt})'$, $\bar{\mathbf{G}}'_w = (\bar{\mathbf{g}}_{w1}, \bar{\mathbf{g}}_{w2}, ..., \bar{\mathbf{g}}_{wT})$, and $\mathbf{X}'_i = (\mathbf{x}_{i1}, \mathbf{x}_{i2}, ..., \mathbf{x}_{iT})$. For fixed N and T,

$$\sum_{t=1}^{T} \mathbf{g}_{w,it}\mathbf{g}'_{w,it} = \begin{pmatrix} \mathbf{X}'_i\mathbf{X}_i & \mathbf{X}'_i\bar{\mathbf{G}}_w \\ \bar{\mathbf{G}}'_w\mathbf{X}_i & \bar{\mathbf{G}}'_w\bar{\mathbf{G}}_w \end{pmatrix}, \tag{2.10}$$

is a non-singular matrix. For a fixed T let $\mathbf{G}'_\mu = (\mathbf{g}_{\mu 1}, \mathbf{g}_{\mu 2}, ..., \mathbf{g}_{\mu T})$, where $\mathbf{g}_{\mu t} = (\mathbf{z}'_t, \boldsymbol{\mu}'_{t,x_1}, f_t)'$, then

$$\frac{1}{T}\sum_{t=1}^{T} \mathbf{g}_{w,it}\mathbf{g}'_{w,it} \overset{p}{\to} \Sigma_{ii,T} = \frac{1}{T}\begin{pmatrix} \mathbf{X}'_i\mathbf{X}_i & \mathbf{X}'_i\mathbf{G}_\mu \\ \mathbf{G}'_\mu\mathbf{X}_i & \mathbf{G}'_\mu\mathbf{G}_\mu \end{pmatrix}, \text{ as } N \to \infty, \tag{2.11}$$

where $\Sigma_{ii,T}$ is a non-stochastic positive definite matrix. Also

$$\frac{1}{T}\sum_{t=1}^{T} \mathbf{g}_{w,it}\mathbf{g}'_{w,it} \overset{p}{\to} \Sigma_{ii} = \begin{pmatrix} \Sigma_{x_i x_i} & \Sigma_{x_i \mu} \\ \Sigma_{\mu x_i} & \Sigma_{\mu\mu} \end{pmatrix}, \text{ as } (N,T) \overset{j}{\to} \infty, \tag{2.12}$$

where $\Sigma_{x_i x_i} = E(\mathbf{x}_{it}\mathbf{x}'_{it})$, $\Sigma_{x_i\mu} = E(\mathbf{x}_{it}\mathbf{g}'_{\mu t})$, $\Sigma_{\mu\mu} = E(\mathbf{g}_{\mu t}\mathbf{g}'_{\mu t})$, and Σ_{ii} is a non-stochastic positive definite matrix.

Assumption 5(c): (identification of $\boldsymbol{\beta}$): For a fixed T

$$\sum_{i=1}^{N} \theta_i \left(\mathbf{X}'_i\mathbf{M}_\mu\mathbf{X}_i\right) \overset{p}{\to} \Psi_T, \text{ as } N \to \infty, \tag{2.13}$$

where the (pooling) weights, θ_i, satisfy the conditions

$$\theta_i = O\left(\frac{1}{N}\right), \text{ and } \sum_{i=1}^{N} |\theta_i| \leq K, \tag{2.14}$$

$$\mathbf{M}_\mu = \mathbf{I}_N - \mathbf{G}_\mu \left(\mathbf{G}'_\mu\mathbf{G}_\mu\right)^{-1}\mathbf{G}'_\mu, \tag{2.15}$$

and Ψ_T is a non-stochastic positive definite matrix. Also

$$\sum_{i=1}^{N}\theta_i\left(\frac{\mathbf{X}_i'\mathbf{M}_\mu\mathbf{X}_i}{T}\right)\xrightarrow{p}\Psi,\text{ as }(N,T)\xrightarrow{j}\infty, \tag{2.16}$$

where Ψ is a non-stochastic positive definite matrix.

Assumption 1(a) is restrictive in two respects: It does not allow for inclusion of lagged y_{it} amongst the regressors, and rules out the inclusion of $I(1)$, integrated of order one, regressors in the model. The latter restriction can be readily relaxed at the expense of further technical complications.[2] Allowing for individual specific dynamics in the model presents new difficulties and is addressed in Pesaran (2002). Assumption 1(b) is quite general and allows for the observed common factors, $\mathbf{z}_t$, to be correlated with $\mathbf{x}_{it}$ and f_t. Also by setting one of the elements of $\mathbf{z}_t$ to unity, fixed effects and non-zero means for the individual-specific regressors, $\mathbf{x}_{it}$, can be accommodated. Assumption 1(c) allows for non-zero correlations between the included regressors and the unobserved common effect, and is sufficiently general for our purposes here. The more general case of multiple factors will be considered in Section 5. Assumption 2(a) is innocuous and corresponds to the fixed effects specification in standard panel data models. Assumption 2(b) imposes stronger restrictions on the distribution of the factor loadings and corresponds to the random effects specification. The random coefficient specification is also assumed for the coefficients of the individual-specific effects, $\boldsymbol{\beta}_i$, in assumption 3.[3] No restrictions are imposed on $\boldsymbol{\alpha}_i$, the coefficients of the common observed effects, apart from assuming that they lie on a compact set. Notice also that the assumption that Ω is a non-negative definite (and not necessarily a positive definite) matrix allows for a sub-set of the slope coefficients to be homogeneous. Assumption 4 allows for heterogeneity of error variances across i, but rules out residual serial correlation, although this part of the assumption can be readily relaxed when T is finite and $x_{it}'s$ are strictly exogenous. Assumption 5(a) is required for identification of the unobserved common effects, f_t, up to a scalar constant, and is met so long as there exists a set of weights, w_i, such that $\sum_{i=1}^{N}w_i\gamma_i\neq 0$. The simple scheme $w_i=1/N$ is an obvious example and yields

$$\sum_{i=1}^{N}|w_i|=1\text{ and }\sum_{i=1}^{N}w_i\gamma_i=N^{-1}\sum_{i=1}^{N}\gamma_i=\bar{\gamma}_N.$$

As we shall see later, in practice, it is possible to check if $\sum_{i=1}^{N}w_i\gamma_i=0$, for a given choice of w_i's. It is also possible to use time-varying weights so long as they are pre-determined and satisfy the conditions of assumption 5(a). Assumptions 5(b) and 5(c) are needed in conjunction with assumption 5(a) for consistent estimation of the parameters of interest, namely $\boldsymbol{\beta}_i$ and their means, $\boldsymbol{\beta}$. Assumption 5(b) is quite general and allows for different degrees of cross section dependence of

[2] The case where $\mathbf{x}_{it}$'s are $I(1)$ whilst z_t is stationary could also be of interest.

[3] For a detailed treatment of the random coefficient model see, for example, Swamy (1970).

the individual specific regressors, $\mathbf{x}_{it}$, as well as between $\mathbf{x}_{it}$, $\bar{\mathbf{x}}_{wt}$, and the common effects, f_t and $\mathbf{z}_t$. Consider the following fairly general model for the individual specific regressors

$$\mathbf{x}_{it} = \boldsymbol{\phi}_{i1} f_t + \Phi_{i2}\mathbf{z}_t + \Phi_{i3}\boldsymbol{\chi}_t + \mathbf{v}_{it}, \ i = 1, 2, ..., N, \tag{2.17}$$

where $\boldsymbol{\chi}_t$ is the vector of common effects specific to $\mathbf{x}_{it}$ (not included in the model for y_{it}), $\mathbf{v}_{it}$ are the idiosyncratic components distributed independently of the common effects and across i with a finite covariance matrix.[4] Averaging these relations across i, using the weights w_i, and assuming that the coefficients $\boldsymbol{\phi}_{i1}$, Φ_{i2}, and Φ_{i3} are bounded in N, we have

$$\bar{\mathbf{x}}_{wt} = \left(\sum_{i=1}^{N} w_i \boldsymbol{\phi}_{i1}\right) f_t + \left(\sum_{i=1}^{N} w_i \Phi_{i2}\right) \mathbf{z}_t + \left(\sum_{i=1}^{N} w_i \Phi_{i3}\right) \boldsymbol{\chi}_t + \sum_{i=1}^{N} w_i \mathbf{v}_{it}. \tag{2.18}$$

It is now easily seen that as $N \to \infty$, $\sum_{i=1}^{N} w_i \mathbf{v}_{it} \overset{q.m.}{\to} 0$, and hence

$$\bar{\mathbf{x}}_{wt} \overset{q.m.}{\to} \boldsymbol{\mu}_{tx} = \bar{\boldsymbol{\phi}}_1 f_t + \bar{\Phi}_2 \mathbf{z}_t + \bar{\Phi}_3 \boldsymbol{\chi}_t, \tag{2.19}$$

where $\bar{\boldsymbol{\phi}}_1$, $\bar{\Phi}_2$, and $\bar{\Phi}_3$ are the limits of $\sum_{i=1}^{N} w_i \boldsymbol{\phi}_{i1}$, $\sum_{i=1}^{N} w_i \Phi_{i2}$, and $\sum_{i=1}^{N} w_i \Phi_{i3}$, respectively as $N \to \infty$. Therefore, a variety of correlation structures (both across i and amongst x_{it}, $\bar{\mathbf{x}}_{wt}$, f_t, and $\mathbf{z}_t$) can be entertained. For example, the Monte Carlo design of the experiments carried out by Coakley, Fuerts and Smith (2002) sets $\bar{\Phi}_2 = \mathbf{0}$, and $\bar{\Phi}_3 = \mathbf{0}$, and therefore implies perfect correlation between $\bar{\mathbf{x}}_{wt}$ and f_t as $N \to \infty$. The analysis of Ahn, Lee and Schmidt (2001) assumes that the individual specific regressors are cross sectionally independent, and in effect imposes zero restrictions on all the elements of $\boldsymbol{\phi}_{i1}$, Φ_{i2}, and Φ_{i3}. But in most applications of interest individual specific regressors are likely to be cross sectionally dependent and a formulation such as (2.17) will be far more widely applicable.

It is also easy to verify that under (2.17), conditions (2.13) and (2.16) of assumption 5(c) hold if it is also assumed that

$$\sum_{i=1}^{N} \theta_i \left(\frac{\mathbf{v}_{it}\mathbf{v}_{it}'}{T}\right)$$

converges in probability to a positive definite matrix; for a fixed T as $N \to \infty$, in the case of (2.13), and as $(N, T) \overset{j}{\to} \infty$, in the case of (2.16).

Finally, it is worth noting that the common feature dynamics across i are captured through the serial correlation structure of the common effects. Other more general individual-specific dynamics can be introduced by relaxing assumption 1 to included lagged values of y_{it}.

[4] It is also possible to extend (2.17) by adding the lagged values of $\mathbf{x}_{it}$ to the right-hand-side variables. Our main results will continue to hold provided the cross section distribution of the eigen values of the dynamic system in $\mathbf{x}_{it}$ satisfy certain restrictions as discussed in Zaffaroni (2001).

3 The Principal Components Estimator

To deal with the residual cross section dependence, Coakley, Fuertes and Smith (2002) propose a principle components estimator by augmenting the regression of y_{it} on $\mathbf{x}_{it}$ with one or more principle components of the estimated residuals, $\hat{u}_{it}$, $i = 1, 2, ..., N$, $t = 1, 2, ..., T$ obtained from the first stage regressions of y_{it} on $\mathbf{x}_{it}$. By means of an example we shall now demonstrate that such an estimator will not be consistent in the general case where f_t and $\bar{\mathbf{x}}_t$ are correlated. Under our assumptions Coakley, Fuertes and Smith (CFS) principal components estimator is consistent either if f_t and $\bar{\mathbf{x}}_t$ are (asymptotically) uncorrelated or if they are (asymptotically) perfectly correlated.

For this purpose we shall focus on the simple case of only one individual-specific regressor ($k = 1$) and assume that all the coefficients of the underlying data generating process are homogeneous across i, namely $\alpha_i = 0$, $\beta_i = \beta$, $\gamma_i = \gamma$, and $\sigma_i^2 = \sigma^2$. This is the set up considered by CFS in the analytical discussion of their estimator. In this case the first principle component is given by $u_t = N^{-1} \sum_{i=1}^{N} u_{it}$. CFS suggest estimating u_t using the pooled estimator of β, given by

$$\hat{\beta}_{PE} = \frac{\sum_{t=1}^{T} \sum_{i=1}^{N} y_{it} x_{it}}{\sum_{t=1}^{T} \sum_{i=1}^{N} x_{it}^2}. \tag{3.1}$$

This yields $\hat{u}_t = N^{-1} \sum_{i=1}^{N} (y_{it} - \hat{\beta}_{PE} x_{it}) = \bar{y}_t - \hat{\beta}_{PE}\ \bar{x}_t$, for $t = 1, 2, ..., T$ which are then used in the augmented OLS regression of y_{it}on x_{it} and $\hat{u}_t$ to obtain the principal components estimate of β, which we denote by $\hat{\beta}_{PC}$.

To examine the asymptotic properties of $\hat{\beta}_{PC}$ as T and $N \to \infty$, using the following vector notations:

$$\begin{aligned}
\mathbf{y}_i &= (y_{i1}, y_{i2}, ..., y_{iT})',\ \mathbf{x}_i = (x_{i1}, x_{i2}, ..., x_{iT})',\ \boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2}, ..., \varepsilon_{iT})' \\
\bar{\mathbf{y}} &= (\bar{y}_1, \bar{y}_2, ..., \bar{y}_T)',\ \bar{\mathbf{x}} = (\bar{x}_1, \bar{x}_2, ..., \bar{x}_T)',\ \bar{\boldsymbol{\varepsilon}} = (\bar{\varepsilon}_1, \bar{\varepsilon}_2, ..., \bar{\varepsilon}_T)' \\
\hat{\mathbf{u}} &= (\hat{u}_1, \hat{u}_2, ..., \hat{u}_T)',\ \mathbf{f} = (f_1, f_2, ..., f_T)',
\end{aligned}$$

we first note that

$$\hat{\beta}_{PC} = \frac{N^{-1} \sum_{i=1}^{N} (\frac{\mathbf{x}_i'\mathbf{y}_i}{T}) - (\frac{\bar{\mathbf{x}}'\hat{\mathbf{u}}}{T})(\frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{T})^{-1}(\frac{\hat{\mathbf{u}}'\bar{\mathbf{y}}}{T})}{D_{NT}}, \tag{3.2}$$

where

$$D_{NT} = N^{-1} \sum_{i=1}^{N} (\frac{\mathbf{x}_i'\mathbf{x}_i}{T}) - (\frac{\bar{\mathbf{x}}'\hat{\mathbf{u}}}{T})(\frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{T})^{-1}(\frac{\hat{\mathbf{u}}'\bar{\mathbf{x}}}{T})$$

In the present simple case

$$\mathbf{y}_i = \beta \mathbf{x}_i + \gamma \mathbf{f} + \boldsymbol{\varepsilon}_i, \tag{3.3}$$

and averaging across i

$$\bar{\mathbf{y}} = \beta \bar{\mathbf{x}} + \gamma \mathbf{f} + \bar{\boldsymbol{\varepsilon}}. \tag{3.4}$$

Using these results in (3.3) we now have

$$\hat{\beta}_{PC}-\beta=\gamma\frac{(\frac{\bar{\mathbf{x}}'\mathbf{f}}{T})-(\frac{\bar{\mathbf{x}}'\hat{\mathbf{u}}}{T})(\frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{T})^{-1}(\frac{\hat{\mathbf{u}}'\mathbf{f}}{T})}{D_{NT}}+\frac{N^{-1}\sum_{i=1}^{N}(\frac{\mathbf{x}_i'\varepsilon_i}{T})-(\frac{\bar{\mathbf{x}}'\hat{\mathbf{u}}}{T})(\frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{T})^{-1}(\frac{\hat{\mathbf{u}}'\bar{\varepsilon}}{T})}{D_{NT}}. \tag{3.5}$$

To derive the probability limit of $\hat{\beta}_{PC}$, as N and $T\to\infty$, we first note that

$$\begin{aligned}\frac{\hat{\mathbf{u}}'\bar{\varepsilon}}{T} &= (\beta-\hat{\beta}_{PE})(\frac{\bar{\mathbf{x}}'\bar{\varepsilon}}{T})+\gamma(\frac{\mathbf{f}'\bar{\varepsilon}}{T})+(\frac{\bar{\varepsilon}'\bar{\varepsilon}}{T}),\\ \frac{\hat{\mathbf{u}}'\bar{\mathbf{x}}}{T} &= (\beta-\hat{\beta}_{PE})(\frac{\bar{\mathbf{x}}'\bar{\mathbf{x}}}{T})+\gamma(\frac{\bar{\mathbf{x}}'\mathbf{f}}{T})+(\frac{\bar{\mathbf{x}}'\bar{\varepsilon}}{T}),\\ \frac{\hat{\mathbf{u}}'\mathbf{f}}{T} &= (\beta-\hat{\beta}_{PE})(\frac{\bar{\mathbf{x}}'\mathbf{f}}{T})+\gamma(\frac{\mathbf{f}'\mathbf{f}}{T})+(\frac{\mathbf{f}'\bar{\varepsilon}}{T}),\end{aligned}$$

and finally

$$\begin{aligned}\frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{T} &= (\beta-\hat{\beta}_{PE})^2(\frac{\bar{\mathbf{x}}'\bar{\mathbf{x}}}{T})+2\gamma(\beta-\hat{\beta}_{PE})(\frac{\bar{\mathbf{x}}'\mathbf{f}}{T})+\gamma^2(\frac{\mathbf{f}'\mathbf{f}}{T})\\ &\quad +(\frac{\bar{\varepsilon}'\bar{\varepsilon}}{T})+2\gamma(\beta-\hat{\beta}_{PE})(\frac{\bar{\mathbf{x}}'\bar{\varepsilon}}{T})+2\gamma(\frac{\mathbf{f}'\bar{\varepsilon}}{T}).\end{aligned}$$

Under the above assumptions $(\frac{\bar{\varepsilon}'\bar{\varepsilon}}{T})$, $(\frac{\bar{\mathbf{x}}'\bar{\varepsilon}}{T})$, $(\frac{\mathbf{f}'\bar{\varepsilon}}{T})$ and $N^{-1}\sum_{i=1}^{N}(\frac{\mathbf{x}_i'\varepsilon_i}{T})$ all converge to zero in probability as N and $T\to\infty$ (in no particular order) and the following probability limits exist and are bounded (see the appendix for proofs)

$$(\frac{\bar{\mathbf{x}}'\bar{\mathbf{x}}}{T})\xrightarrow{p}\sigma_{\bar{x}}^2\geq 0,\ (\frac{\bar{\mathbf{x}}'\mathbf{f}}{T})\xrightarrow{p}\sigma_{\bar{x}f},\ (\frac{\mathbf{f}'\mathbf{f}}{T})\xrightarrow{p}\sigma_f^2>0,$$

and

$$\frac{1}{N}\sum_{i=1}^{N}(\frac{\mathbf{x}_i'\mathbf{x}_i}{T})\xrightarrow{p}\lim_{N\to\infty}\left(\frac{1}{N}\sum_{i=1}^{N}\sigma_{ix}^2\right)=\sigma_x^2>0.$$

Also using (3.1)

$$\beta-\hat{\beta}_{PE}\xrightarrow{p}-\gamma\left(\frac{\sigma_{\bar{x}f}}{\sigma_x^2}\right).$$

Substituting these probability limits in (3.5) and after some algebra we have

$$\hat{\beta}_{PC}-\beta\xrightarrow{p}\frac{\gamma\left(\sigma_{\bar{x}f}/\sigma_x^2\right)\left(\sigma_f^2\sigma_{\bar{x}}^2-\sigma_{\bar{x}f}^2\right)}{\sigma_x^2\sigma_f^2-\sigma_{\bar{x}f}^2\left[\sigma_{\bar{x}}^4/\sigma_x^4-3\sigma_{\bar{x}}^2/\sigma_x^2+3\right]}. \tag{3.6}$$

Therefore, in the presence of common effects ($\gamma\neq 0$) the CFS's principal components estimator is consistent only under the two extremes of zero correlation between the common factor and the cross-section average of the included regressor, namely if $\sigma_{\bar{x}f}=0$, and when the common factor and the cross section average of the included regressor are perfectly correlated, namely if $\sigma_{\bar{x}f}^2=\sigma_f^2\sigma_{\bar{x}}^2$. This result also explains CFS's Monte Carlo simulations and the small sample evidence that they seem to provide in support of their proposed estimator. The processes used to generate f_t and x_{it} are given by

$$\begin{aligned} f_t &= 0.9\ f_{t-1} + \varepsilon_{ft}, \\ x_{it} &= d_{it} + \lambda_i f_t, \\ d_{it} &= 0.9\ d_{i,t-1} + \varepsilon_{di,t}, \end{aligned}$$

and the shocks ε_{ft} and $\varepsilon_{di,t}$ are IID draws from the normal distribution. It is now easily seen that

$$\bar{x}_t = \bar{d}_t + \bar{\lambda} f_t,$$

where $\bar{d}_t$ and $\bar{\lambda}$ are the cross section means of d_{it} and λ_i, respectively. Also

$$\bar{d}_t = 0.9\ \bar{d}_{t-1} + \bar{\varepsilon}_{dt},$$

and since the shocks, $\varepsilon_{di,t}$, are IID it then readily follows that $Var(\bar{\varepsilon}_{dt}) \to 0$ and hence $Var(\bar{d}_t) \to 0$ for each t as $N \to \infty$. Therefore, $\bar{x}_t$ and f_t will become perfectly correlated if N is sufficiently large. Perfect correlation between $\bar{x}_t$ and f_t can be avoided, for example, by allowing a sufficient degree of dependence across d_{it} so that $Var(\bar{d}_t)$ does not go to zero.

4 A General Approach to Estimation of Panels with Common Effects

The main difficulty with the CFS's estimator lies in the fact that it makes use of an inconsistent estimator of $\boldsymbol{\beta}$ to obtain the principal components which are then used as proxies for the unobserved common effects. To see how this problem can be overcome initially we work with the homogeneous case, a single individual-specific variable and no observed common effects, namely we set $\boldsymbol{\alpha}_i = \mathbf{0}$, $\beta_i = \beta$, $\gamma_i = \gamma$, and $\sigma_i^2 = \sigma^2$, for all i. For these parameters and averaging (2.1) using the weights $w_i = 1/N$, we have

$$\bar{y}_t = \beta \bar{x}_t + \gamma f_t + \bar{\varepsilon}_t. \tag{4.1}$$

Under assumption 4, $Var(\bar{\varepsilon}_t) = \sigma^2/N$, and $\bar{\varepsilon}_t$ converges to zero in root mean square error. Therefore, so long as $\gamma \neq 0$ and for N sufficiently large, up to a scalar constant, f_t converges in probability to $\bar{y}_t - \beta \bar{x}_t$.[5] It is therefore possible to identify and estimate the unobserved common effects, f_t, only if a consistent estimate of β is available. Hence the direct use of z_t for the purpose of consistently

[5]The validity of condition $\gamma \neq 0$ can be checked in practice if N and T are sufficiently large. Note that under $\gamma = 0$, and N sufficiently large $\bar{y}_t \approx \beta \bar{x}_t$ and the regression of $\bar{y}_t$ on $\bar{x}_t$, the so called aggregate function associated with the underlying "micro" equations (2.1), must yield a perfect fit. As we shall see below this result is quite general and extends to the heterogeneous slope case with multiple factors.

estimating β will not be a fruitful strategy. However, for consistent estimation of β a consistent estimate of f_t is not necessarily needed. It proves adequate to use $\bar{y}_t$ and $\bar{x}_t$ separately as they together form a sufficient basis for the consistent estimation of f_t. This observation suggests running a regression of y_{it} on x_{it} augmented with the cross section averages of the dependent variable and the regressors. We shall refer to this estimator as the **correlated common effect estimator**. In the present simple case this estimator is given by

$$\hat{\beta}_{CC} = \frac{N^{-1}\sum_{i=1}^{N}(\frac{\mathbf{x}_i'\mathbf{y}_i}{T}) - (\frac{\bar{\mathbf{x}}'\bar{\mathbf{H}}}{T})(\frac{\bar{\mathbf{H}}'\bar{\mathbf{H}}}{T})^{-1}(\frac{\bar{\mathbf{H}}'\bar{\mathbf{y}}}{T})}{D_{NT}}, \tag{4.2}$$

where $\bar{\mathbf{H}} = (\bar{\mathbf{y}}, \bar{\mathbf{x}})$ and

$$D_{NT} = N^{-1}\sum_{i=1}^{N}\left(\frac{\mathbf{x}_i'\mathbf{x}_i}{T}\right) - \left(\frac{\bar{\mathbf{x}}'\bar{\mathbf{H}}}{T}\right)\left(\frac{\bar{\mathbf{H}}'\bar{\mathbf{H}}}{T}\right)^{-1}\left(\frac{\bar{\mathbf{H}}'\bar{\mathbf{x}}}{T}\right). \tag{4.3}$$

However, since $\bar{\mathbf{x}}$ is contained in $\bar{\mathbf{H}}$, then $\bar{\mathbf{H}}(\bar{\mathbf{H}}'\bar{\mathbf{H}})^{-1}\bar{\mathbf{H}}\bar{\mathbf{x}} = \bar{\mathbf{x}}$, and the expression for $\hat{\beta}_{CC}$ simplifies further to

$$\hat{\beta}_{CC} = \frac{N^{-1}\sum_{i=1}^{N}(\frac{\mathbf{x}_i'\mathbf{y}_i}{T}) - (\frac{\bar{\mathbf{x}}'\bar{\mathbf{y}}}{T})}{N^{-1}\sum_{i=1}^{N}(\frac{\mathbf{x}_i'\mathbf{x}_i}{T}) - (\frac{\bar{\mathbf{x}}'\bar{\mathbf{x}}}{T})}. \tag{4.4}$$

Using (3.3) and (4.1) now yields

$$\hat{\beta}_{CC} - \beta = \frac{N^{-1}\sum_{i=1}^{N}(\frac{\mathbf{x}_i'\varepsilon_i}{T}) - \frac{\bar{\mathbf{x}}'\bar{\varepsilon}}{T}}{N^{-1}\sum_{i=1}^{N}(\frac{\mathbf{x}_i'\mathbf{x}_i}{T}) - (\frac{\bar{\mathbf{x}}'\bar{\mathbf{x}}}{T})},$$

which is free of the nuisance parameter, γ. It is now easily seen that as $N \to \infty$ (irrespective of whether T is fixed or tends to infinity)

$$\hat{\beta}_{CC} \xrightarrow{p} \beta,$$

provided for each t as $N \to \infty$ we have

$$\frac{1}{N}\sum_{i=1}^{N} x_{it}^2 \xrightarrow{p} \sigma_{xt}^2 > \sigma_{\bar{x}_t}^2 > 0, \tag{4.5}$$

where $\sigma_{\bar{x}_t}^2 = \lim_{N\to\infty} E(\bar{x}_t^2)$. Then for a fixed T and as $N \to \infty$

$$\hat{\beta}_{CC} - \beta \xrightarrow{p} \frac{p\lim_{N\to\infty}\left[N^{-1}\sum_{i=1}^{N}\mathbf{x}_i'\varepsilon_i\right] - p\lim_{N\to\infty}(\bar{\mathbf{x}}'\bar{\varepsilon})}{\sum_{t=1}^{T}(\sigma_{xt}^2 - \sigma_{\bar{x}_t}^2)}.$$

But it is easily seen that

$$E\left(\bar{\mathbf{x}}'\bar{\varepsilon}\right) = 0$$

$$Var(\bar{\mathbf{x}}'\bar{\boldsymbol{\varepsilon}}) = \frac{\sigma^2}{N} E(\bar{\mathbf{x}}'\bar{\mathbf{x}}).$$

Similarly,

$$Var\left[N^{-1}\sum_{i=1}^{N}\mathbf{x}_i'\varepsilon_i\right] = \frac{\sigma^2}{N}\left[\frac{\sum_{i=1}^{N}E(\mathbf{x}_i'\mathbf{x}_i)}{N}\right].$$

Therefore for a fixed T, as $N \to \infty$, $N^{-1}\sum_{i=1}^{N}\mathbf{x}_i'\varepsilon_i$ and $\bar{\mathbf{x}}'\bar{\boldsymbol{\varepsilon}}$ tend to zero in probability and $\hat{\beta}_{CC} \xrightarrow{p} \beta$. Also as $T \to \infty$, $\hat{\beta}_{CC} \xrightarrow{p} \beta$ for any fixed $N \geq 2$, so long as

$$\left(\frac{1}{N}\sum_{i=1}^{N}\sigma_{ix}^2\right) > \sigma_{\bar{x}}^2. \tag{4.6}$$

For $N = 2$ this estimator reduces to the OLS estimator of the slope in the regression of $y_{1t} - y_{2t}$ on $x_{1t} - x_{2t}, t = 1, 2, ..., T$. Conditions (4.5) and (4.6) ensure that the individual-specific regressors exhibit adequate time and cross section variations. The consistency property of $\hat{\beta}_{CC}$ continues to hold under the joint asymptotics when both T and $N \to \infty$.

Finally, it is worth noting that under $\gamma_i = \gamma$, the CCE estimator, (4.4), is in fact the same as the familiar de-meaned regression estimator given by:

$$\hat{\beta}_{de-meaned} = \frac{\sum_{i=1}^{N}(\mathbf{x}_i - \bar{\mathbf{x}})'(\mathbf{y}_i - \bar{\mathbf{y}})}{\sum_{i=1}^{N}(\mathbf{x}_i - \bar{\mathbf{x}})'(\mathbf{x}_i - \bar{\mathbf{x}})}.$$

The algebraic equivalence of these two estimators in the case where all the coefficients are homogeneous is re-assuring but not surprising. What would be of interest is to see if the proposed estimator continues to be valid in the heterogeneous case where γ_i differ across i, particularly considering that the de-meaned regression estimator fails to produce a consistent estimator in this case even under $\beta_i = \beta$. Note that

$$\mathbf{y}_i - \bar{\mathbf{y}} = \beta(\mathbf{x}_i - \bar{\mathbf{x}}) + (\gamma_i - \bar{\gamma})\mathbf{f} + (\boldsymbol{\varepsilon}_i - \bar{\boldsymbol{\varepsilon}}),$$

where $\bar{\gamma} = N^{-1}\sum_{i=1}^{N}\gamma_i$. Hence

$$\hat{\beta}_{de-meaned} - \beta = \frac{\sum_{i=1}^{N}\gamma_i(\mathbf{x}_i - \bar{\mathbf{x}})'\mathbf{f}}{\sum_{i=1}^{N}(\mathbf{x}_i - \bar{\mathbf{x}})'(\mathbf{x}_i - \bar{\mathbf{x}})} + \frac{\sum_{i=1}^{N}\mathbf{x}_i'(\boldsymbol{\varepsilon}_i - \bar{\boldsymbol{\varepsilon}})}{\sum_{i=1}^{N}(\mathbf{x}_i - \bar{\mathbf{x}})'(\mathbf{x}_i - \bar{\mathbf{x}})}.$$

The first term on the right hand side of the above expression is identically equal to zero only if $\gamma_i = \gamma$, but in general where the unobserved common effects are correlated with the included regressors this term does not vanish even for N and T sufficiently large.

In the general heterogeneous case two different, but related, estimation problems must be addressed, namely estimation of the individual coefficients, $\boldsymbol{\alpha}_i$, $\boldsymbol{\beta}_i$, σ_i^2 and possibly γ_i, and the average slope coefficients, $\boldsymbol{\beta}$, defined by (2.5).

4.1 Panels with Heterogenous Effects: Individual Specific Coefficients

Focusing on the individual-specific regressions, consistent estimates of $\boldsymbol{\alpha}_i$, $\boldsymbol{\beta}_i$, σ_i^2 can be obtained by running the OLS regression of y_{it} on $\mathbf{x}_{it}$, $\mathbf{z}_t$, $\bar{y}_{wt}$, and $\bar{\mathbf{x}}_{1wt}$ where[6]

$$\bar{y}_{wt} = \sum_{j=1}^{N} w_j y_{jt}, \;\; \bar{\mathbf{x}}_{1wt} = \sum_{j=1}^{N} w_j \mathbf{x}_{1jt}. \tag{4.7}$$

In principle the weights used in the construction of the aggregates, $\bar{y}_{wt}$ and $\bar{\mathbf{x}}_{wt}$, could be individual-specific, namely one could even use

$$\bar{y}_{w_it} = \sum_{j=1}^{N} w_{ij} y_{jt}, \;\; \bar{\mathbf{x}}_{w_it} = \sum_{j=1}^{N} w_{ij} \mathbf{x}_{jt}, \tag{4.8}$$

with $w_{ii} = 0$. As we shall see later the optimal choice of these weights will depend on the unknown parameters, γ_j and σ_j^2, $j = 1, 2, ..., N$. But for consistent estimation it is only required that the chosen weights satisfy the conditions of assumption 5(a), namely that $\sum_{j=1}^{N} w_{ij}^2 \to 0$ as $N \to \infty$, and that $\sum_{j=1}^{N} w_{ij}\gamma_j \neq 0$. As noted earlier, a simple and obvious choice is $w_{ij} = 1/N$ for all i and j. But before a set of weights is selected the validity of $\sum_{j=1}^{N} w_{ij}\gamma_j \neq 0$ needs to be checked. Fortunately this is possible even though this condition depends on the unknown factor loadings, γ_i. Consider the weights $\{w_j\}$ and assume that $\sum_{j=1}^{N} w_j^2 \to 0$, as $N \to \infty$. Also to simplify the exposition without loss of generality we assume that all the individual specific regressors are sufficiently cross sectionally dependent such that for each t, namely $\bar{\mathbf{x}}_{wt} \overset{q.m.}{\to} \boldsymbol{\mu}_{\bar{x}t} \neq \mathbf{0}$.[7] Then under assumptions 1(a), 3, 4, 5(a) and 5(b), for each t we have[8]

$$\bar{y}_{wt} - \left(\mathbf{z}_t'\bar{\boldsymbol{\alpha}}_w + \bar{\mathbf{x}}_{wt}'\boldsymbol{\beta} + \bar{\gamma}_w f_t\right) \overset{q.m.}{\to} 0, \text{ as } N \to \infty, \tag{4.9}$$

where

$$\bar{\gamma}_w = \sum_{j=1}^{N} w_j \gamma_j, \;\; \bar{\boldsymbol{\alpha}}_w = \sum_{j=1}^{N} w_j \boldsymbol{\alpha}_j. \tag{4.10}$$

Therefore, the OLS regression of $\bar{y}_{wt}$ on $\mathbf{z}_t$ and $\bar{\mathbf{x}}_{wt}$ must fit perfectly if $\bar{\gamma}_w = 0$, and N is sufficiently large. In what follows we suppose that this is not the case and the variables $\mathbf{z}_t$, $\mathbf{x}_{it}$, $\bar{\mathbf{x}}_{wt}$ and $\bar{y}_{wt}$, are not perfectly correlated. Under these assumptions the correlated common effects estimator of $\boldsymbol{\beta}_i$ exists and is given by the OLS estimate of $\mathbf{b}_i$ in the augmented regression

$$\mathbf{y}_i = \mathbf{Z}\mathbf{a}_i + \mathbf{X}_i\mathbf{b}_i + \bar{\mathbf{X}}_w\mathbf{c}_{i1} + \bar{\mathbf{y}}_w c_{i2} + \mathbf{e}_i, \tag{4.11}$$

[6] Recall from assumption 5(b) that $\mathbf{x}_{1it}$ is the sub-set of the individual specific regressors assumed to be sufficiently cross sectionally dependent.

[7] See assumption 5b for more detail.

[8] See Proposition A.1 in the Appendix.

where $\mathbf{y}_i$ is the $T \times 1$ vector observations on y_{it}, $\mathbf{Z}$ and $\mathbf{X}_i$ are the $T \times k_z$, and $T \times k_x$ matrices of observations on $\mathbf{z}_t$ and $\mathbf{x}_{it}$, respectively, and $\bar{\mathbf{X}}_w$ and $\bar{\mathbf{y}}_w$ are $T \times k_f$ and $T \times 1$ observation matrices on the aggregates $\bar{\mathbf{x}}_{wt}$ and $\bar{y}_{wt}$, respectively.[9] Throughout we suppose that $T > k_z + 2k_x + 1$,and $(\mathbf{Z}, \mathbf{X}_i, \bar{\mathbf{X}}_w, \bar{\mathbf{y}}_w)$ is a full column rank matrix. Using familiar results from partitioned regressions the CCE of $\boldsymbol{\beta}_i$ can be written as

$$\hat{\mathbf{b}}_i = (\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{X}_i)^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{y}_i, \tag{4.12}$$

where

$$\bar{\mathbf{M}}_w = \mathbf{I}_T - \bar{\mathbf{H}}_w(\bar{\mathbf{H}}_w'\bar{\mathbf{H}}_w)^{-1}\bar{\mathbf{H}}_w', \tag{4.13}$$

and

$$\bar{\mathbf{H}}_w = (\mathbf{Z}, \bar{\mathbf{X}}_w, \bar{\mathbf{y}}_w). \tag{4.14}$$

To establish the consistency of this estimator we first write (2.1) and (2.2) in matrix notations as

$$\mathbf{y}_i = \mathbf{Z}\boldsymbol{\alpha}_i + \mathbf{X}_i\boldsymbol{\beta}_i + \gamma_i\mathbf{f} + \boldsymbol{\varepsilon}_i, \tag{4.15}$$

together with its associated aggregate form

$$\bar{\mathbf{y}}_w = \mathbf{Z}\bar{\boldsymbol{\alpha}}_w + \bar{\mathbf{X}}_w\boldsymbol{\beta} + \bar{\gamma}_w\mathbf{f} + \bar{\boldsymbol{\xi}}_w, \tag{4.16}$$

where $\bar{\boldsymbol{\alpha}}_w$ and $\bar{\gamma}_w$ are already defined in (4.10) and

$$\begin{aligned} \bar{\mathbf{X}}_w &= \sum_{j=1}^{N} w_j\mathbf{X}_j,\ \bar{\boldsymbol{\xi}}_w = \sum_{j=1}^{N} w_j\boldsymbol{\xi}_j, \\ \boldsymbol{\xi}_j &= \boldsymbol{\varepsilon}_j + \mathbf{X}_j\boldsymbol{\upsilon}_j, \end{aligned} \tag{4.17}$$

with $\boldsymbol{\upsilon}_j$ defined by (2.5). Under random coefficients, the composite error term, $\boldsymbol{\xi}_i$, are independently distributed across j even when the regressors, $\mathbf{X}_j$, are cross-sectionally dependent. Using (4.15) in (4.12) we have

$$\begin{aligned} \hat{\mathbf{b}}_i - \boldsymbol{\beta}_i &= \gamma_i\left(\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{X}_i}{T}\right)^{-1}\left(\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{f}}{T}\right) \\ &\quad + \left(\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{X}_i}{T}\right)^{-1}\left(\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\boldsymbol{\varepsilon}_i}{T}\right). \end{aligned} \tag{4.18}$$

Also since by assumption $\bar{\gamma}_w \neq 0$, using (4.16) we have

$$\mathbf{f} = \bar{\gamma}_w^{-1}\left(\bar{\mathbf{y}}_w - \mathbf{Z}\bar{\boldsymbol{\alpha}}_{\mathbf{w}} - \bar{\mathbf{X}}_w\boldsymbol{\beta} - \bar{\boldsymbol{\xi}}_w\right). \tag{4.19}$$

[9] In the case where some of the individual specific regressors are cross sectionally independent, cross-section averages of these regressors should not be included in the augmented regression.

But by construction $\bar{\mathbf{M}}_w \left(\bar{\mathbf{y}}_w - \mathbf{Z}\bar{\boldsymbol{\alpha}}_\mathbf{w} - \bar{\mathbf{X}}_w\boldsymbol{\beta}\right) = \mathbf{0}$. Hence

$$\bar{\mathbf{M}}_w\mathbf{f} = -\bar{\gamma}_w^{-1}\bar{\boldsymbol{\xi}}_w, \tag{4.20}$$

and

$$\begin{aligned}\hat{\mathbf{b}}_i - \boldsymbol{\beta}_i &= -\left(\frac{\gamma_i}{\bar{\gamma}_w}\right)\left(\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{X}_i}{T}\right)^{-1}\left(\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\bar{\boldsymbol{\xi}}_w}{T}\right) \\ &\quad + \left(\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{X}_i}{T}\right)^{-1}\left(\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\boldsymbol{\varepsilon}_i}{T}\right).\end{aligned} \tag{4.21}$$

It is now easily seen that for a finite N the CCE estimator, $\hat{\mathbf{b}}_i$, will depend on the unknown factor loadings and in general will be biased, even if $T \to \infty$. To see this note that

$$\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\bar{\boldsymbol{\xi}}_w}{T} = \frac{\sum_{t=1}^T \mathbf{x}_{it}\bar{\xi}_{wt}}{T} - \left(\frac{\sum_{t=1}^T \mathbf{x}_{it}\bar{\mathbf{h}}_{wt}'}{T}\right)\left(\frac{\sum_{t=1}^T \bar{\mathbf{h}}_{wt}\bar{\mathbf{h}}_{wt}'}{T}\right)^{-1}\left(\frac{\sum_{t=1}^T \bar{\mathbf{h}}_{wt}\bar{\xi}_{wt}}{T}\right),$$

where $\bar{\mathbf{h}}_{wt} = (\mathbf{z}_t', \bar{\mathbf{x}}_{wt}', \bar{y}_{wt})'$. Using (A.19)

$$\bar{\mathbf{h}}_{wt} = \mathbf{A}_w\bar{\mathbf{g}}_{wt} + \bar{\boldsymbol{\nu}}_{wt}, \tag{4.22}$$

where $\bar{\mathbf{g}}_{wt} = (\mathbf{z}_t', \bar{\mathbf{x}}_{wt}', f_t)'$,

$$\mathbf{A}_w = \begin{pmatrix} \mathbf{I}_{k_z} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{k_x} & \mathbf{0} \\ \bar{\boldsymbol{\alpha}}_w' & \boldsymbol{\beta}' & \bar{\gamma}_w \end{pmatrix}, \text{ and } \bar{\boldsymbol{\nu}}_{wt} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \bar{\xi}_{wt} \end{pmatrix}, \tag{4.23}$$

Under our assumptions for a fixed N

$$\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\boldsymbol{\varepsilon}_i}{T} \xrightarrow{p} \mathbf{0}, \text{ and } \frac{\sum_{t=1}^T \mathbf{x}_{it}\bar{\xi}_{wt}}{T} \xrightarrow{p} \mathbf{0}, \text{ as } T \to \infty. \tag{4.24}$$

But

$$\frac{\sum_{t=1}^T \bar{\mathbf{h}}_{wt}\bar{\xi}_{wt}}{T} = \mathbf{A}_w\left(\frac{\sum_{t=1}^T \bar{\mathbf{g}}_{wt}\bar{\xi}_{wt}}{T}\right) + \frac{\sum_{t=1}^T \bar{\boldsymbol{\nu}}_{wt}\bar{\xi}_{wt}}{T},$$

and

$$\frac{\sum_{t=1}^T \bar{\mathbf{h}}_{wt}\bar{\xi}_{wt}}{T} \xrightarrow{p} \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ p\lim_{T\to\infty}\left[\frac{\sum_{t=1}^T \bar{\xi}_{wt}^2}{T}\right] \end{pmatrix},$$

which converges to a non-zero value if N is fixed.[10] For the CCE estimator to be consistent it is therefore necessary that N is sufficiently large so that the dependence of $\hat{\mathbf{b}}_i - \boldsymbol{\beta}_i$ on the factor

[10]In the simple case where $\boldsymbol{\beta}_i = \boldsymbol{\beta}$,

$$p\lim_{T\to\infty}\left(\frac{\sum_{t=1}^T \bar{\xi}_{wt}^2}{T}\right) = p\lim_{T\to\infty}\left(\frac{\sum_{t=1}^T \bar{\varepsilon}_{wt}^2}{T}\right) = \frac{\sum_{i=1}^N \sigma_i^2}{N^2} \neq 0, \text{ for a fixed } N.$$

loadings disappears in the limit. The following theorem provides a formal statement of this result and the associated asymptotic distributions.

Theorem 4.1 *Consider the panel data model (2.1) and (2.2) and suppose that assumptions 1, 2(a), 3, 4, 5(a) and 5(b) hold.*

(a) - (N-asymptotic) The correlated common effects estimator, $\hat{\mathbf{b}}_i$*, defined by (4.12) is unbiased for a fixed* $T > k_z + 2k_x + 1$ *and* $N \to \infty$*, in the sense that* $lim_{N\to\infty} E\left(\hat{\mathbf{b}}_i\right) = \boldsymbol{\beta}_i$*. Under the additional assumption that* $\varepsilon_{it} \sim N(0, \sigma_i^2)$*,*

$$\hat{\mathbf{b}}_i - \boldsymbol{\beta}_i \xrightarrow{d} N(\mathbf{0}, \Sigma_{T,b_i}), \tag{4.25}$$

as $N \to \infty$*, where*

$$\Sigma_{T,b_i} = \sigma_i^2 \Psi_{iT}^{-1}, \quad \Psi_{iT} = \mathbf{X}_i' \mathbf{M}_\mu \mathbf{X}_i, \tag{4.26}$$

$$\mathbf{M}_\mu = \mathbf{I}_T - \mathbf{G}_\mu (\mathbf{G}_\mu' \mathbf{G}_\mu)^{-1} \mathbf{G}_\mu', \tag{4.27}$$

$\mathbf{G}_\mu = (\mathbf{g}_{\mu 1}, \mathbf{g}_{\mu 2}, ..., \mathbf{g}_{\mu T})$*,* $\mathbf{g}_{\mu t} = (\mathbf{z}_t', \, \boldsymbol{\mu}_{tx}', \, f_t)'$*, and* $\boldsymbol{\mu}_{tx} = p\lim_{N\to\infty}(\bar{\mathbf{x}}_{wt})$*.*

(b) - (Joint asymptotics) As $(N,T) \xrightarrow{j} \infty$ *(in no particular order),* $\hat{\mathbf{b}}_i$ *is a consistent estimator of* $\boldsymbol{\beta}_i$*. If it is further assumed that* $\sqrt{T}/N \to 0$ *(or if* $T/N \to \kappa$*, where* κ *is a fixed non-zero constant) as* $(N,T) \xrightarrow{j} \infty$*, then*

$$\sqrt{T}\left(\hat{\mathbf{b}}_i - \boldsymbol{\beta}_i\right) \xrightarrow{d} N(\mathbf{0}, \Sigma_{b_i}), \tag{4.28}$$

where

$$\Sigma_{b_i} = \sigma_i^2 \Psi_i^{-1}, \; \Psi_i = \Sigma_{x_i x_i} - \Sigma_{x_i \mu} \Sigma_{\mu\mu}^{-1} \Sigma_{\mu x_i}. \tag{4.29}$$

An asymptotically unbiased estimator of Σ_{T,b_i} is given by (as $N \to \infty$ for fixed $T > k_z + 2k_x + 1$)

$$\hat{\Sigma}_{T,b_i} = \hat{\sigma}_i^2 \left(\mathbf{X}_i' \bar{\mathbf{M}}_w \mathbf{X}_i\right)^{-1}, \tag{4.30}$$

where

$$\hat{\sigma}_i^2 = \frac{\left(\mathbf{y}_i - \mathbf{X}_i \hat{\mathbf{b}}_i\right)' \bar{\mathbf{M}}_w \left(\mathbf{y}_i - \mathbf{X}_i \hat{\mathbf{b}}_i\right)}{T - (k_z + 2k_x + 1)}. \tag{4.31}$$

In the case where $(N,T) \xrightarrow{j} \infty$, a consistent estimator of $\Sigma_{b_i}$is given by

$$\hat{\Sigma}_{b_i} = \hat{\sigma}_i^2 \left(\frac{\mathbf{X}_i' \bar{\mathbf{M}}_w \mathbf{X}_i}{T}\right)^{-1}, \tag{4.32}$$

For a proof see the Appendix.

4.2 Pooled Correlated Common Effects Estimators

In the case where the parameters of interest are the cross-section means of slope coefficients $\boldsymbol{\beta}_i$, namely $\boldsymbol{\beta}$ defined by (2.5), one possibility would be to use the mean group (MG) estimator proposed by Pesaran and Smith (1995), which is a simple average of the individual estimators, $\hat{\mathbf{b}}_i$, given by (4.12).

$$\hat{\mathbf{b}}_{MG} = N^{-1} \sum_{i=1}^{N} \hat{\mathbf{b}}_i. \tag{4.33}$$

As an alternative one could also consider the Swamy's Random Coefficient (RC) estimator defined by the weighted average of the individual estimates with the weights being inversely proportional to the individual variances (see, for example, Swamy (1970)):

$$\hat{\mathbf{b}}_{RC} = \sum_{i=1}^{N} \Theta_i \hat{\mathbf{b}}_i, \tag{4.34}$$

where

$$\Theta_i = \left\{ \sum_{j=1}^{N} \left[\hat{\Sigma}_{T,b_j} + \hat{\Omega} \right]^{-1} \right\}^{-1} \left[\hat{\Sigma}_{T,b_i} + \hat{\Omega} \right]^{-1}, \tag{4.35}$$

$\hat{\Sigma}_{T,b_j}$ is given by (4.30) and $\hat{\Omega}$ is a consistent estimator of Ω, the variance of $\boldsymbol{\beta}_i$ defined by (2.5). A comparative analysis of the MG and the RC estimators in the context of dynamic panel data models without unobserved common effects is provided in Hsiao, Pesaran and Tahmiscioglu (1999). It is shown that for N and T sufficiently large both of these estimators are consistent and asymptotically equivalent. These results continue to apply in the more general setting of this paper. In particular, the MGE is asymptotically unbiased as $N \to \infty$, for a fixed T, and will be consistent as $(N,T) \xrightarrow{j} \infty$. Furthermore, as $(N,T) \xrightarrow{j} \infty$ with $T/N \to \kappa$, where κ is a fixed non-zero constant, the asymptotic distribution of $\hat{\mathbf{b}}_{MG}$ (or $\hat{\mathbf{b}}_{RC}$) is given by

$$\sqrt{N} \left(\hat{\mathbf{b}}_{MG} - \boldsymbol{\beta} \right) \xrightarrow{d} N(\mathbf{0}, \Omega), \tag{4.36}$$

where Ω is consistently estimated by

$$\hat{\Omega} = \frac{1}{N} \sum_{i=1}^{N} \left(\mathbf{b}_i - \hat{\mathbf{b}}_{MG} \right) \left(\mathbf{b}_i - \hat{\mathbf{b}}_{MG} \right)'. \tag{4.37}$$

However, these estimators are likely to have poor small sample properties particularly if T is small.[11] Efficiency gains from pooling of observations over the cross section units can be achieved

[11]Monte Carlo evidence on the small sample properties of the MG and RC estimators are provided in Hsiao, Pesaran and Tahmiscioglu (1999), where it is shown that Swamy estimator can also be viewed as an empirical Bayes estimator.

when the individual slope coefficients, $\boldsymbol{\beta}_i$, are the same. In what follows we developed a pooled estimator of $\boldsymbol{\beta}$ that assumes (possibly incorrectly) that $\boldsymbol{\beta}_i = \boldsymbol{\beta}$, and $\sigma_i^2 = \sigma^2$, although it allows the slope coefficients of the common effects (whether observed or not) to differ across i. Such a pooled estimator of $\boldsymbol{\beta}$ is given by

$$\hat{\mathbf{b}} = \left(\sum_{i=1}^{N} \theta_i \mathbf{X}_i' \bar{\mathbf{M}}_w \mathbf{X}_i \right)^{-1} \sum_{i=1}^{N} \theta_i \mathbf{X}_i' \bar{\mathbf{M}}_w \mathbf{y}_i. \tag{4.38}$$

Typically, the (pooling) weights θ_i is set equal to $1/N$, although in the general case where σ_i^2 differ across i as we shall it will be optimal to set $\theta_i = \sigma_i^{-2} / \sum_{j=1}^{N} \sigma_j^{-2}$. However, in practice where σ_i^2 is unknown the efficiency gain from using an estimate of σ_i^2 is likely to be limited particularly when T is small. Nevertheless, to maintain a reasonable level of generality in what follows we allow θ_i to differ across i but treat them as non-stochastic constants satisfying the conditions in (2.14).

Although, it is not necessary for the (pooling) weights, θ_i, to be the same as the weights, w_i, used in construction of the aggregates, $\bar{y}_{wt}$ and $\bar{\mathbf{x}}_{wt}$; for the invariance of the asymptotic distribution of $\hat{\mathbf{b}}$ to the factor loadings, γ_i, we must have

$$\left(\sum_{i=1}^{N} \theta_i^2 \right)^{-1/2} \left(\sum_{i=1}^{N} \theta_i \mathbf{X}_i \right)' \bar{\mathbf{M}}_w \bar{\boldsymbol{\xi}}_w \xrightarrow{p} \mathbf{0}, \tag{4.39}$$

and this is satisfied for all N if $\theta_i = w_i$. Under this restriction $\sum_{i=1}^{N} \theta_i \mathbf{X}_i = \bar{\mathbf{X}}_w$ and it follows immediately that $\left(\sum_{i=1}^{N} \theta_i \mathbf{X}_i \right)' \bar{\mathbf{M}}_w = \mathbf{0}$.

We refer to $\hat{\mathbf{b}}$ as the "pooled correlated common effects" (PCCE) estimator. It can also be viewed as a "generalized" fixed effects estimator, in the sense that the PCCE estimator reduces to the fixed effects estimator in the case where $\mathbf{z}_t = 1$, $f_t = 0$, and $\theta_i = \theta$ for all t and i. Within our framework the fixed effects can be viewed as time-invariant common effects with heterogeneous slopes.

As before, we investigate the asymptotic properties of this estimator as $N \to \infty$, both when T is fixed and when $(N, T) \xrightarrow{j} \infty$. Using (4.15) and (4.20) we have

$$\hat{\mathbf{b}} - \boldsymbol{\beta} = \left(\sum_{i=1}^{N} \theta_i \mathbf{X}_i' \bar{\mathbf{M}}_w \mathbf{X}_i \right)^{-1} \sum_{i=1}^{N} \left[\theta_i \mathbf{X}_i' \bar{\mathbf{M}}_w \boldsymbol{\xi}_i - \gamma_i \theta_i \left(\mathbf{X}_i' \bar{\mathbf{M}}_w \bar{\boldsymbol{\xi}}_w \right) \right], \tag{4.40}$$

where without loss of generality we have set $\bar{\gamma}_w = \sum_{j=1}^{N} w_j \gamma_j = 1$. For a fixed N, even if $T \to \infty$, the distribution of $\hat{\mathbf{b}}$ will depend on the unknown factor loadings and will be biased. In general, we need N to be sufficiently large, although T could be fixed as $N \to \infty$. Also whilst it is possible to show that $\hat{\mathbf{b}}$ is a consistent estimator of $\boldsymbol{\beta}$ with a minimal set of restrictions on the factor loadings, γ_i, for asymptotic normality of the pooled estimator we shall require the more restrictive assumption that γ_i's follow a random coefficient model as set out in assumption 2(b). The following theorem provides a formal statement of our main results.

Theorem 4.2 *Consider the panel data model (2.1) and (2.2) and suppose that assumptions 1, 2(a), 3, 4, 5(a) and 5(c) hold.*

(a) - (N-asymptotic) The pooled correlated common effects estimator, $\hat{\mathbf{b}}$, *defined by (4.38) is a consistent estimator of* $\boldsymbol{\beta}$ *for a fixed* $T > k_z + 2k_x + 1$ *and as* $N \to \infty$. *Under the additional assumption 2(b) and for* $\theta_i = w_i$

$$\left(\sum_{i=1}^{N} w_i^2\right)^{-1/2} \left(\hat{\mathbf{b}} - \boldsymbol{\beta}\right) \xrightarrow{d} N(\mathbf{0}, \Sigma_{b,T}) \tag{4.41}$$

where

$$\Sigma_{\mathbf{b},T} = \Psi_T^{-1}\mathbf{R}_T\Psi_T^{-1}, \tag{4.42}$$

$$\Psi_T = plim_{N\to\infty}\left[\sum_{i=1}^{N} w_i \Psi_{iT}\right], \tag{4.43}$$

$$\mathbf{R}_T = plim_{N\to\infty}\left\{\sum_{i=1}^{N} \tilde{w}_i^2 \left[\sigma_i^2 \Psi_{iT} + \Psi_{iT}\Omega\Psi_{iT}\right]\right\}, \tag{4.44}$$

$$\tilde{w}_i = \frac{w_i}{\sqrt{\sum_{j=1}^{N} w_j^2}}, \tag{4.45}$$

and Ψ_{iT} *is defined by (4.26).*

(b) - (Joint asymptotics) As $(N,T) \xrightarrow{j} \infty$ *(in no particular order)* $\hat{\mathbf{b}}$ *is a consistent estimator of* $\boldsymbol{\beta}$. *Under the additional assumption 2(b), if* $\theta_i = w_i$ *and* $\sqrt{T}/N \to 0$ *(or* $T/N \to \kappa$, *where* κ *is a fixed non-zero constant) as* $(N,T) \xrightarrow{j} \infty$, *then*

$$\left(\sum_{i=1}^{N} w_i^2\right)^{-1/2} \left(\hat{\mathbf{b}} - \boldsymbol{\beta}\right) \xrightarrow{d} N(\mathbf{0}, \Sigma_b), \tag{4.46}$$

where

$$\Sigma_b = \Psi^{-1}\mathbf{R}\Psi^{-1}, \tag{4.47}$$

$$\Psi = \lim_{N\to\infty}\left(\sum_{i=1}^{N} w_i \Psi_i\right), \ \mathbf{R} = \lim_{N\to\infty}\left(\sum_{i=1}^{N} \tilde{w}_i^2 \Psi_i \Omega \Psi_i\right), \tag{4.48}$$

and Ψ_i *is defined by (4.29).*

Consistent estimators of $\Sigma_{b,T}$ and Σ_b can be obtained using

$$\hat{\Psi}_T = \sum_{i=1}^{N} w_i \hat{\Psi}_{iT},\ \hat{\mathbf{R}}_T = \sum_{i=1}^{N} \tilde{w}_i^2 \left[\hat{\sigma}_i^2 \hat{\Psi}_{iT} + \hat{\Psi}_{iT}\hat{\Omega}\hat{\Psi}_{iT}\right],$$

$$\hat{\Psi} = \frac{1}{T}\sum_{i=1}^{N} w_i \hat{\Psi}_{iT},\ \hat{\mathbf{R}}_T = \frac{1}{T^2}\sum_{i=1}^{N} \tilde{w}_i^2 \hat{\Psi}_{iT}\hat{\Omega}\hat{\Psi}_{iT},$$

where $\hat{\Psi}_{iT} = \mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{X}_i$, $\hat{\Omega}$ is given by (4.37) and $\hat{\sigma}_i^2$ by (4.31).

As with the Swamy and the MG estimators the above pooled estimator is likely to perform well for N and T to be sufficiently large if the slopes, $\boldsymbol{\beta}_i$, differ across i. In fact it is easy to show that $\hat{\mathbf{b}}$ is asymptotically dominated by $\hat{\mathbf{b}}_{MG}$ (or $\hat{\mathbf{b}}_{RC}$) under slope heterogeneity and for $\theta_i = 1/N$. In this case we have

$$\begin{aligned} AVar\left(\sqrt{N}\hat{\mathbf{b}}\right) - AVar\left(\sqrt{N}\hat{\mathbf{b}}_{MG}\right) &= \Psi^{-1}\mathbf{R}\Psi^{-1} - \Omega \\ &= \Psi^{-1}\left(\mathbf{R} - \Psi\Omega\Psi\right)\Psi^{-1}, \end{aligned}$$

and since Ψ is a positive definite matrix we need only consider $\mathbf{R} - \Psi\Omega\Psi$. But

$$\begin{aligned} \left(\mathbf{R} - \Psi\Omega\Psi\right) &= \lim_{N\to\infty}\left[\frac{1}{N}\sum_{i=1}^{N}\Psi_i\Omega\Psi_i - \left(\frac{1}{N}\sum_{i=1}^{N}\Psi_i\right)\Omega\left(\frac{1}{N}\sum_{i=1}^{N}\Psi_i\right)\right] \\ &= \lim_{N\to\infty}\left[\frac{1}{N}\sum_{i=1}^{N}\left(\Psi_i - \bar{\Psi}\right)\Omega\left(\Psi_i - \bar{\Psi}\right)\right], \end{aligned}$$

where $\bar{\Psi} = \frac{1}{N}\sum_{i=1}^{N}\Psi_i$, which establishes the desired result given that Ω is a positive definite matrix. Therefore, the pooled estimator, $\hat{\mathbf{b}}$, can (asymptotically) dominate the mean group (or RC) estimator only when $\boldsymbol{\beta}_i$'s are reasonably homogeneous across i. It is also interesting that in the present application where the regressors are strictly exogenous the heterogeneity of the effects of the common factors (observed or not) do not affect the asymptotic distribution of the pooled estimator.

However, the above result does not hold in the case where $\boldsymbol{\beta}_i$'s are homogeneous, namely when $\Omega = \mathbf{0}$.[12] In this case the rate of convergence of $\hat{\mathbf{b}}$ to $\boldsymbol{\beta}$ will also depend on T, and its asymptotic covariance matrix, as $(N,T) \xrightarrow{j} \infty$, is no longer given by (4.47). The asymptotic results for the homogeneous case is summarized in the following theorem.

Theorem 4.3 *Consider the panel data model (2.1) and (2.2) and suppose that assumptions 1, 2(b), (4), 5(a), and 5(c) hold and $\boldsymbol{\beta}_i = \boldsymbol{\beta}$ for all i.*

(a) - (N-asymptotic) For a fixed $T > k_z + 2k_x + 1$, and as $N \to \infty$

$$\left(\sum_{i=1}^{N} w_i^2\right)^{-1/2}\left(\hat{\mathbf{b}} - \boldsymbol{\beta}\right) \xrightarrow{d} N(\mathbf{0}, \Sigma_{b,T}) \tag{4.49}$$

[12]Here we do not consider intermediate cases where a sub-set of $\boldsymbol{\beta}_i$ could be homogenous.

where

$$\Sigma_{b,T} = \Psi_T^{-1} \dot{\mathbf{R}}_T \Psi_T^{-1}, \tag{4.50}$$

$$\Psi_T = plim_{N\to\infty} \left[\sum_{i=1}^{N} w_i \Psi_{iT} \right], \tag{4.51}$$

$$\dot{\mathbf{R}}_T = plim_{N\to\infty} \left\{ \sum_{i=1}^{N} \tilde{w}_i^2 \sigma_i^2 \Psi_{iT} \right\}. \tag{4.52}$$

(b) - (Joint asymptotics) As $(N,T) \xrightarrow{j} \infty$ *such that* $\sqrt{T}/N \to 0$ *(or* $T/N \to \kappa$*, with* κ *being a fixed non-zero constant)*

$$\left(\frac{\sum_{i=1}^{N} w_i^2}{T} \right)^{-1/2} \left(\hat{\mathbf{b}} - \boldsymbol{\beta} \right) \xrightarrow{d} N(\mathbf{0}, \Sigma_b), \tag{4.53}$$

where

$$\Sigma_b = \Psi^{-1} \dot{\mathbf{R}} \Psi^{-1}, \tag{4.54}$$

$$\Psi = \lim_{N\to\infty} \left(\sum_{i=1}^{N} w_i \Psi_i \right), \ \dot{\mathbf{R}} = \lim_{N\to\infty} \left(\sum_{i=1}^{N} \tilde{w}_i^2 \sigma_i^2 \Psi_i \right). \tag{4.55}$$

The asymptotic variance matrix of $\hat{\mathbf{b}}$ under $\boldsymbol{\beta}_i = \boldsymbol{\beta}$ is given by

$$AVar\left(\hat{\mathbf{b}}\right) = \left(\frac{\sum_{i=1}^{N} w_i^2}{T} \right) \left(\Psi^{-1} \dot{\mathbf{R}} \Psi^{-1} \right),$$

which upon using (4.45) reduces to

$$AVar\left(\hat{\mathbf{b}}\right) = \frac{1}{T} \left(\sum_{i=1}^{N} w_i \Psi_i \right)^{-1} \left(\sum_{i=1}^{N} w_i^2 \sigma_i^2 \Psi_i \right) \left(\sum_{i=1}^{N} w_i \Psi_i \right)^{-1}. \tag{4.56}$$

Alternative consistent estimators of $AVar\left(\hat{\mathbf{b}}\right)$ can be obtained depending on the size of T relative to N. When T is of the same order of magnitude as N, one could use

$$\widehat{AVar}\left(\hat{\mathbf{b}}\right) = \frac{1}{T} \left(\sum_{i=1}^{N} w_i \frac{\mathbf{X}_i' \bar{\mathbf{M}}_w \mathbf{X}_i}{T} \right)^{-1} \left(\sum_{i=1}^{N} w_i^2 \hat{\sigma}_i^2 \frac{\mathbf{X}_i' \bar{\mathbf{M}}_w \mathbf{X}_i}{T} \right) \left(\sum_{i=1}^{N} w_i \frac{\mathbf{X}_i' \bar{\mathbf{M}}_w \mathbf{X}_i}{T} \right)^{-1}, \tag{4.57}$$

where $\hat{\sigma}_i^2$ is given by (4.31). On the other hand if T is small one could follow Arellano (1987) and use the robust estimator of the asymptotic variance of $\hat{\mathbf{b}}$ given by

$$\widehat{AVar}\left(\hat{\mathbf{b}}\right) = \frac{1}{T} \left(\frac{\sum_{i=1}^{N} w_i \hat{\mathbf{X}}_i' \hat{\mathbf{X}}_i}{T} \right)^{-1} \left(\frac{\sum_{i=1}^{N} w_i^2 \hat{\mathbf{X}}_i' \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i' \hat{\mathbf{X}}_i}{T} \right) \left(\frac{\sum_{i=1}^{N} w_i \hat{\mathbf{X}}_i' \hat{\mathbf{X}}_i}{T} \right)^{-1}, \tag{4.58}$$

where

$$\hat{\mathbf{X}}_i = \bar{\mathbf{M}}_w \mathbf{X}_i, \text{ and } \hat{\mathbf{u}}_i = \bar{\mathbf{M}}_w \left(\mathbf{y}_i - \mathbf{X}_i \hat{\mathbf{b}}\right). \tag{4.59}$$

This estimator is valid in the presence of error variance heterogeneity and serial correlation in ε_{it}, $t = 1, 2, ..., T$, provided T is small relative to N. To allow only for error variance heterogeneity the middle term in (4.58) must be replaced by

$$\frac{\sum_{t=1}^{T} \sum_{i=1}^{N} w_i^2 \hat{u}_{it}^2 \hat{\mathbf{x}}_{it} \hat{\mathbf{x}}_{it}'}{T},$$

where $\hat{\mathbf{x}}_{it}$ is the t^{th} column of $\hat{\mathbf{X}}_i'$, and $\hat{u}_{it}$ is the t^{th} element of $\hat{\mathbf{u}}_i$.

4.3 Determination of Optimal Weights

Our results hold for all values of $\{w_i,\ i = 1, 2, ..., N\}$ that satisfy the three conditions set out in assumption 5(a). But it is clear that these conditions do not uniquely determine the weights and the issue of how to choose w_i's optimally presents itself. One possible approach would be to determine the weights such that the asymptotic variance of the estimators of interest are minimized (in a suitable sense) subject to the conditions of assumption 5(a) being met. For the individual coefficients, $\hat{\mathbf{b}}_i$, this amounts to minimizing $\Sigma_{T,b_i} = \sigma_i^2 \Psi_{iT}^{-1}$, given by (4.26), which is the same as maximizing $\Psi_{iT} = \mathbf{X}_i' \mathbf{M}_\mu \mathbf{X}_i$ with respect to w_i's. However, it is easily seen that Ψ_{iT} does not depend on w_i's. To see this note that Ψ_{iT} can be viewed as sample correlation matrix of residuals obtained from regressions of the columns of $\mathbf{X}_i$ on $\mathbf{G}_\mu = (\mathbf{g}_{\mu 1}, \mathbf{g}_{\mu 2}, ..., \mathbf{g}_{\mu T})$, where $\mathbf{g}_{\mu t} = (\mathbf{z}_t',\ \boldsymbol{\mu}_{tx}',\ f_t)'$. But under (2.17), which presents a fairly general specification of the $\mathbf{x}_{it}$ process, we have $\boldsymbol{\mu}_{tx} = \bar{\boldsymbol{\phi}}_1 f_t + \bar{\Phi}_2 \mathbf{z}_t + \bar{\Phi}_3 \boldsymbol{\chi}_t$, and $\mathbf{G}_\mu$ depends on w_i's, only through the coefficients $\bar{\boldsymbol{\phi}}_1$, $\bar{\Phi}_2$, and $\bar{\Phi}_3$. Therefore, $\mathbf{X}_i' \mathbf{M}_\mu \mathbf{X}_i$ will be identical to the sample correlation matrix of the residuals from the regressions of $\mathbf{X}_i$ on f_t, $\mathbf{z}_t$, and $\boldsymbol{\chi}_t$, and hence will be invariant to the choice of w_i's. The same also applies to Ψ_i, which is the probability limit of $T^{-1}\left(\mathbf{X}_i' \mathbf{M}_\mu \mathbf{X}_i\right)$, as $T \to \infty$.

Consider now the asymptotic variance of the pooled estimator, $\hat{\mathbf{b}}$, given by (4.56). Noting that Ψ_i's are invariant to the choice of $\{w_i\}$, it is then easily established that subject to (2.7), $AVar\left(\hat{\mathbf{b}}\right)$ is minimized with θ_i set at

$$\theta_i^* = \frac{\sigma_i^{-2}}{\sum_{j=1}^{N} \sigma_j^{-2}}. \tag{4.60}$$

First, $\sum_{i=1}^{N} |\theta_i^*| = 1$ and under assumption 4 it is easily verified that $\theta_i^* = O\left(\frac{1}{N}\right)$.[13] Also $AVar\left(\hat{\mathbf{b}}\right)$ evaluated at θ_i^* yields

$$AVar\left(\hat{\mathbf{b}}(\theta^*)\right) = \frac{1}{T}\left(\sum_{i=1}^{N} \sigma_i^{-2} \Psi_i\right)^{-1}. \tag{4.61}$$

[13] The third condition in (2.7), namely $\sum_{j=1}^{N} \gamma_j \sigma_j^{-2} \neq 0$, is an identification restriction which is assumed *a priori*.

But noting that Ψ_i is a positive definite matrix we can write

$$T\left[AVar\left(\hat{\mathbf{b}}(w^*)\right)^{-1} - AVar\left(\hat{\mathbf{b}}\right)^{-1}\right]$$

$$= \left(\sum_{i=1}^{N}\mathbf{A}_i\mathbf{A}_i'\right) - \left(\sum_{i=1}^{N}\mathbf{A}_i\mathbf{B}_i'\right)\left(\sum_{i=1}^{N}\mathbf{B}_i\mathbf{B}_i'\right)^{-1}\left(\sum_{i=1}^{N}\mathbf{B}_i\mathbf{A}_i'\right) \geq \mathbf{0},$$

where

$$\mathbf{A}_i = \sigma_i^{-1}\Psi_i^{1/2},\ \mathbf{B}_i = \theta_i\sigma_i\Psi_i^{1/2}.$$

This now establishes that $\left[AVar\left(\hat{\mathbf{b}}(\theta^*)\right)^{-1} - AVar\left(\hat{\mathbf{b}}\right)^{-1}\right]$ is a non-negative definite matrix, with $\{\theta_i^*\}$ providing an optimal choice in the sense that $AVar\left(\hat{\mathbf{b}}(\theta^*)\right) \leq AVar\left(\hat{\mathbf{b}}\right)$.

Not surprisingly the pooled estimator computed using θ_i^* reduces to the generalized least squares estimator

$$\hat{\mathbf{b}}(\theta^*) = \left(\sum_{i=1}^{N}\sigma_i^{-2}\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{X}_i\right)^{-1}\sum_{i=1}^{N}\sigma_i^{-2}\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{y}_i, \tag{4.62}$$

with its feasible counterpart obtained by replacing σ_i^2 with the estimates, $\hat{\sigma}_i^2$, given by (4.31) and computed using an initial consistent estimator of $\boldsymbol{\beta}$ based on (say) $\theta_i = 1/N$. Recall, however, for the pooled estimator to remain asymptotically valid the weights used for the construction of the aggregates should, in general, be the same as the ones used in the formation of the pooled estimator.[14]

5 A Multi-Factor Generalization

The analysis of the previous section can be readily generalized to allow for more than one unobserved common factor. Consider the following generalization of (2.2)

$$u_{it} = \boldsymbol{\gamma}_i'\mathbf{f}_t + \varepsilon_{it}, \tag{5.1}$$

where $\mathbf{f}_t$ is now an $m \times 1$ vector of unobserved common effects and $\boldsymbol{\gamma}_i$ is the associated individual specific vector of factor loadings. Suppose that there exists m different sets of weights, w_{li}, for $l = 1, 2, ..., m$ and $i = 1, 2, ..., N$, such that the following conditions (which are a generalization of the conditions in (2.7)) are met:

$$\text{(i): } w_{li} = O\left(\frac{1}{N}\right),\ \text{(ii): } \sum_{i=1}^{N}|w_{li}| < K, \quad \text{for } l = 1, 2, ..., m, \tag{5.2}$$

[14] It would be valid to use different sets of weights if the individual specific regressors, $\mathbf{x}_{it}$, were asymptotically perfectly correlated with the common effects, $\mathbf{z}_t$ and $\mathbf{f}_t$.

and (iii): the $m \times m$ matrix

$$\bar{\Gamma}' = (\bar{\boldsymbol{\gamma}}_1, \bar{\boldsymbol{\gamma}}_2, ..., \bar{\boldsymbol{\gamma}}_m), \tag{5.3}$$

is non-singular for all N (including as $N \to \infty$) where

$$\bar{\boldsymbol{\gamma}}_l = \sum_{j=1}^{N} w_{lj} \boldsymbol{\gamma}_j. \tag{5.4}$$

Using the weights, w_{li}, the cross section aggregation of (2.1) now yields (for each t)

$$\bar{y}_{lt} = \bar{\boldsymbol{\alpha}}_l' \mathbf{z}_t + \bar{\mathbf{x}}_{lt}' \boldsymbol{\beta} + \bar{\boldsymbol{\gamma}}_l' \mathbf{f}_t + \bar{\xi}_{lt}, \;\; l = 1, 2, ..., m, \tag{5.5}$$

where

$$\begin{aligned} \bar{y}_{lt} &= \sum_{j=1}^{N} w_{lj} y_{jt}, \; \bar{\mathbf{x}}_{lt} = \sum_{j=1}^{N} w_{lj} \mathbf{x}_{jt}, \; \bar{\xi}_{lt} = \sum_{j=1}^{N} w_{lj} \xi_{jt}, \\ \xi_{jt} &= \varepsilon_{jt} + \mathbf{x}_{jt}' \boldsymbol{\upsilon}_j, \text{ and } \bar{\boldsymbol{\alpha}}_l = \sum_{j=1}^{N} w_{lj} \boldsymbol{\alpha}_j. \end{aligned} \tag{5.6}$$

As in the single factor case, it is easily seen that as $N \to \infty$ (also see proposition A.1),

$$\bar{y}_{lt} - \left(\bar{\boldsymbol{\alpha}}_l' \mathbf{z}_t + \bar{\mathbf{x}}_{lt}' \boldsymbol{\beta} + \bar{\boldsymbol{\gamma}}_l' \mathbf{f}_t \right) \overset{q.m.}{\to} 0 \text{ for } l = 1, 2, ..., m, \tag{5.7}$$

and provided $\bar{\Gamma}$ is non-singular, then for each t and as $N \to \infty$, the m unknown factors, $\mathbf{f}_t$, can be proxied perfectly by linear combinations of the observables, $\bar{y}_{lt}$, $\bar{\mathbf{x}}_{lt}$, $l = 1, 2, ..., m$, and $\mathbf{z}_t$. More specifically, for each t we have

$$\mathbf{f}_t - \bar{\Gamma}^{-1} \left(\bar{\mathbf{y}}_t - \bar{\mathbf{A}} \mathbf{z}_t - \bar{\mathbb{X}}_t \boldsymbol{\beta} \right) \overset{q.m.}{\to} \mathbf{0}, \tag{5.8}$$

where $\bar{\mathbf{y}}_t = (\bar{y}_{1t}, \bar{y}_{2t}, ..., \bar{y}_{mt})'$, $\bar{\mathbf{A}}' = (\bar{\boldsymbol{\alpha}}_1, \bar{\boldsymbol{\alpha}}_2, ..., \bar{\boldsymbol{\alpha}}_m)$, and $\bar{\mathbb{X}}_t' = (\bar{\mathbf{x}}_{1t}, \bar{\mathbf{x}}_{2t}, ..., \bar{\mathbf{x}}_{mt})$.[15] Therefore, in this more general setting the correlated common effects estimator for the individual- specific slopes, $\boldsymbol{\beta}_i$, is given by

$$\hat{\mathbf{b}}_i = (\mathbf{X}_i' \bar{\mathbf{M}} \mathbf{X}_i)^{-1} \mathbf{X}_i' \bar{\mathbf{M}} \mathbf{y}_i, \tag{5.9}$$

where

$$\bar{\mathbf{M}} = \mathbf{I}_T - \bar{\mathbf{H}} (\bar{\mathbf{H}}' \bar{\mathbf{H}})^{-1} \bar{\mathbf{H}}', \tag{5.10}$$

$$\bar{\mathbf{H}} = (\mathbf{Z}, \bar{\mathbf{X}}, \bar{\mathbf{Y}}), \tag{5.11}$$

[15] This result also shows that the unobserved factors can be identified only up to a non-singular transformation and without loss of generality $\bar{\Gamma}$ can be set to an identity matrix of order m.

$\bar{\mathbf{Y}} = (\bar{\mathbf{y}}_1, \bar{\mathbf{y}}_2, ..., \bar{\mathbf{y}}_T)'$, $\bar{\mathbf{X}} = \left(\bar{\mathbf{X}}_1, \bar{\mathbf{X}}_2, ..., \bar{\mathbf{X}}_m\right)$, and $\bar{\mathbf{X}}_l = (\bar{\mathbf{x}}_{l1}, \bar{\mathbf{x}}_{l2}, ..., \bar{\mathbf{x}}_{lT})'$. Note that $\bar{\mathbf{Y}}$ is $T \times m$, and $\bar{\mathbf{X}}$ is $T \times m\ k_x$, and for $\hat{\mathbf{b}}_i$ to exist it is necessary that $T > (m+1)k_x + k_z + m$. These are direct generalizations of (4.12), (4.13) and (4.14). Similarly, the pooled correlated common effects estimator in this case is given by[16]

$$\hat{\mathbf{b}} = \left(\sum_{i=1}^{N} \bar{w}_i \mathbf{X}_i' \bar{\mathbf{M}} \mathbf{X}_i\right)^{-1} \sum_{i=1}^{N} \bar{w}_i \mathbf{X}_i' \bar{\mathbf{M}} \mathbf{y}_i, \tag{5.12}$$

where

$$\bar{w}_i = \frac{\sum_{l=1}^{m} w_{li}}{m}. \tag{5.13}$$

To allow for error variance heterogeneity at the estimation stage we could also set

$$\bar{w}_i^* = \frac{\sigma_i^{-2}}{\sum_{j=1}^{N} \sigma_j^{-2}}. \tag{5.14}$$

Following similar lines of proof, it can be shown that theorems 4.1 and 4.2 developed for the residual one-factor model equally applies to the current more general case, provided m is fixed as N and $T \to \infty$.

As in the one-factor case there are numerous ways that the weights $\{w_{li},\ l = 1, 2, ..., m;\ i = 1, 2, ..., N\}$ can be set. One possibility would be to partition the N individual observations into m different groups of (approximately) equal size. This can be done, for example, by a randomization procedure, or can be based on *a priori* criteria, such as geographical or social groupings. The same weights could then be applied to members of a given group and zero to the observations that lie outside the group. More specifically, let $N = \sum_{i=1}^{m} N_i$, with N_1, $N_2, ..., N_m$ being of similar orders of magnitude. Then set

$$\begin{aligned} w_{1i} &= 1/N_1, \text{ for } i = 1, 2, ..., N_1 \\ &= 0, \text{ otherwise,} \end{aligned}$$

$$\begin{aligned} w_{2i} &= 1/N_2, \text{ for } i = N_1 + 1, N_1 + 2, ..., N_1 + N_2, \\ &= 0, \text{ otherwise,} \end{aligned}$$

and so on, with the last set of weights defined by

$$\begin{aligned} w_{mi} &= 1/N_m, \text{ for } i = \sum_{i=1}^{m-1} N_i + 1, \sum_{i=1}^{m-1} N_i + 2, ..., \sum_{i=1}^{m} N_i, \\ &= 0, \text{ otherwise.} \end{aligned}$$

It is easy to verify that these weights satisfy the conditions in (5.2), provided that as $N \to \infty$, $N_i/N \to \kappa_i$, where κ_i is a fixed non-zero constant.

[16] For asymptotic (as $N \to \infty$) invariance of the distribution of $\hat{\mathbf{b}}$ to the nuisance parameters, $\boldsymbol{\gamma}_i$, we must have $\left(\sum_{i=1}^{N} \bar{w}_i \mathbf{X}_i\right)' \bar{\mathbf{M}} = \mathbf{0}$, which is clearly satisfied for the choice of $\bar{w}_i$ given by (5.13).

6 Concluding Remarks

This paper provides a simple procedure for estimation of panel data models subject to cross section dependence when the cross section dimension (N) of the panel is sufficiently large. The asymptotic theory required for estimation and inference is developed under fairly general conditions both when the time dimension (T) is fixed and when $T \to \infty$. Conditions under which the proposed correlated common effects estimators are consistent and asymptotically normal are provided. Further extensions and generalizations are, however, clearly desirable.

It is assumed that m, the number of unobserved factors is known. In principle it should be possible to adapt the analysis of Bai and Ng (2002) to estimate m. One possibility would be to directly apply the Bai-Ng procedure to the residuals

$$\hat{\mathbf{u}}_i = \bar{\mathbf{M}}\left(\mathbf{y}_i - \mathbf{X}_i\hat{\mathbf{b}}_i\right), \text{ or } \hat{\mathbf{u}}_i = \bar{\mathbf{M}}\left(\mathbf{y}_i - \mathbf{X}_i\hat{\mathbf{b}}\right).$$

Under our assumptions these residuals provide consistent estimates of u_{it} in the multi-factor model (5.1) and could be used as "observed data" to obtain estimates of the factors $\mathbf{f}_t$ (up to a non-singular transformation). It is a reasonable to expect these factor estimates (denoted by $\hat{\mathbf{f}}_t$) to be consistent, and could justify the application of Bai-Ng procedure. The factor estimates can also be used directly as (generated) regressors in the regression equation

$$y_{it} = \mathbf{z}_t'\mathbf{a}_i + \mathbf{x}_{it}'\mathbf{b}_i + \hat{\mathbf{f}}_t'\mathbf{c}_i + \zeta_{it},$$

to evaluate the joint statistical significance of the unobserved factors. Since factors can be identified only up to a non-singular transformation, statistical analyses of the effects of individual factors will be problematic.

Also it is desirable to see if the results of this paper carry over to the case where lagged values of y_{it} are allowed to be included amongst the individual-specific regressors. The regression model (2.1) allows for dynamics only through the general dynamics of the aggregate effects and the fact that these effects could have differential impacts on different groups. This is restrictive and its relaxation is clearly important for a wider applicability of the approach advanced in this paper.

Another important extension is to multi-variate panel data models such as Panel Vector Autoregressions (PVAR) of the type discussed, for example, in Binder, Hsiao and Pesaran (2002).

Finally, small sample properties of the proposed estimators need to be evaluated by Monte Carlo experiments and compared with alternative estimators when available.

These further developments are beyond the scope of the present paper and will be the subject of separate studies.

Appendix

Lemma A.1 *Under assumptions 1(a), (3), (4), and 5(a) we have*

$$E(\bar{\xi}_{wt}) = 0, \tag{A.1}$$

$$E(\bar{\xi}_{wt}\varepsilon_{it}) = w_i\sigma_i^2 = O\left(w_i\right) = O\left(\frac{1}{N}\right), \tag{A.2}$$

$$E(\bar{\xi}_{wt}^2) = \sum_{j=1}^{N} w_j^2\sigma_j^2 + \sum_{j=1}^{N} w_j^2 E(\mathbf{x}_{jt}'\Omega\mathbf{x}_{jt}) = O\left(\sum_{j=1}^{N} w_j^2\right) = O\left(\frac{1}{N}\right), \tag{A.3}$$

$$Var(\bar{\xi}_{wt}\varepsilon_{it}) = w_i^2\left[E(\varepsilon_{it}^4) + \sigma_i^2 E(\mathbf{x}_{it}'\Omega\mathbf{x}_{it}) - \sigma_i^4\right] = O\left(w_i^2\right) = O\left(\frac{1}{N^2}\right), \tag{A.4}$$

and

$$Var(\bar{\xi}_{wt}^2) = O\left[\left(\sum_{j=1}^{N} w_j^2\right)^2\right] = O\left(\frac{1}{N^2}\right), \tag{A.5}$$

where $\bar{\xi}_{wt}$ is the average error term defined by

$$\bar{\xi}_{wt} = \sum_{j=1}^{N} w_j\varepsilon_{jt} + \sum_{j=1}^{N} w_j\mathbf{x}_{jt}'\boldsymbol{v}_j. \tag{A.6}$$

Also

$$Var\left(\frac{1}{T}\sum_{t=1}^{T}\bar{\xi}_{wt}\right) = O\left(\frac{\sum_{j=1}^{N} w_j^2}{T}\right) = O\left(\frac{1}{T\ N}\right), \tag{A.7}$$

$$Var\left(\frac{1}{T}\sum_{t=1}^{T}\bar{\xi}_{wt}\varepsilon_{it}\right) = \frac{w_i^2}{T}\left\{\frac{1}{T}\sum_{t=1}^{T}\left[E(\varepsilon_{it}^4) - \sigma_i^4 + \sigma_i^2 E(\mathbf{x}_{it}'\Omega\mathbf{x}_{it})\right]\right\} = O\left(\frac{w_i^2}{T}\right), \tag{A.8}$$

$$Var\left(\frac{1}{T}\sum_{t=1}^{T}\bar{\xi}_{wt}^2\right) \leq Var(\bar{\xi}_{wt}^2) = O\left(\frac{1}{N^2}\right), \tag{A.9}$$

$$E\left(\frac{1}{T}\sum_{t=1}^{T}\bar{\xi}_{wt}^2\right) = O\left(\sum_{j=1}^{N} w_j^2\right) = O\left(\frac{1}{N}\right), \tag{A.10}$$

$$Var\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_{it}\bar{\xi}_{wt}\right)=O\left(\frac{1}{N\ T}\right), \tag{A.11}$$

$$Var\left(\frac{1}{T}\sum_{t=1}^{T}\bar{\mathbf{g}}_{wt}\bar{\xi}_{wt}\right)=O\left(\frac{1}{N\ T}\right), \tag{A.12}$$

where $\bar{\mathbf{g}}_{wt}=(\mathbf{z}_t',\bar{\mathbf{x}}_{wt}',f_t)'$.

Proof. First using (A.6) it is easily seen that under assumptions 1(a), 3 and 4, for each t,

$$E(\bar{\xi}_{wt})=0.$$

Also

$$\begin{aligned} E(\varepsilon_{it}\bar{\xi}_{wt}) &= E\left[\varepsilon_{it}\left(\sum_{j=1}^{N}w_j\varepsilon_{jt}+\sum_{j=1}^{N}w_j\mathbf{x}_{jt}'\boldsymbol{\upsilon}_j\right)\right] \\ &= w_i\sigma_i^2, \end{aligned}$$

and (A.2) follows since $\sigma_i^2\leq K$.

Consider now

$$Var(\bar{\xi}_{wt})=E(\bar{\xi}_{wt}^2)=\sum_{j=1}^{N}w_j^2\sigma_j^2+\sum_{j=1}^{N}w_j^2E(\mathbf{x}_{jt}'\Omega\mathbf{x}_{jt}). \tag{A.13}$$

But $\mathbf{x}_{jt}'\Omega\mathbf{x}_{jt}\leq\lambda_{\max}(\Omega)\left(\mathbf{x}_{jt}'\mathbf{x}_{jt}\right)$, where $\lambda_{\max}(\Omega)$ is the maximum eigen value of Ω. Hence

$$Var(\bar{\xi}_{wt})\leq\left(\sum_{j=1}^{N}w_j^2\sigma_j^2\right)+\lambda_{\max}(\Omega)\sum_{j=1}^{N}w_j^2E(\mathbf{x}_{jt}'\mathbf{x}_{jt}),$$

Also under assumptions 1(a) and 4, σ_j^2 and $E(\mathbf{x}_{jt}'\mathbf{x}_{jt})$ are bounded in N for all j and t, and hence

$$E(\bar{\xi}_{wt}^2)=Var(\bar{\xi}_{wt})\leq K\sum_{j=1}^{N}w_j^2, \tag{A.14}$$

for some finite positive constant, K. Consequently, under assumption 1(a), 3, 4 and 5(a) for each t

$$Var(\bar{\xi}_{wt})=O\left(\sum_{j=1}^{N}w_j^2\right)=O(\frac{1}{N}). \tag{A.15}$$

Similarly,

$$Var(\bar{\xi}_{wt}\varepsilon_{it})=E(\bar{\xi}_{wt}^2\varepsilon_{it}^2)-\left[E(\bar{\xi}_{wt}\varepsilon_{it})\right]^2.$$

But

$$
\begin{aligned}
E(\bar{\xi}_{wt}^2 \varepsilon_{it}^2) &= \sum_{k,l} w_k w_l E\left(\varepsilon_{it}^2 \xi_{kt} \xi_{lt}\right) \\
&= w_i^2 E\left(\varepsilon_{it}^2 \xi_{it}^2\right)
\end{aligned}
$$

where

$$\xi_{it} = \varepsilon_{it} + \mathbf{x}_{it}' \boldsymbol{\upsilon}_i.$$

Also it is easily seen that

$$E(\bar{\xi}_{wt}^2 \varepsilon_{it}^2) = w_i^2 \left[E\left(\varepsilon_{it}^4\right) + \sigma_i^2 E(\mathbf{x}_{it}' \Omega \mathbf{x}_{it})\right].$$

Hence using this result in conjunction with (A.2) we have

$$Var(\bar{\xi}_{wt} \varepsilon_{it}) = w_i^2 \left[E\left(\varepsilon_{it}^4\right) + \sigma_i^2 E(\mathbf{x}_{it}' \Omega \mathbf{x}_{it}) - \sigma_i^4\right].$$

Under our assumptions the term after w_i^2 is bounded in N, and hence

$$Var(\bar{\xi}_{wt} \varepsilon_{it}) = O\left(w_i^2\right).$$

Consider now $Var(\bar{\xi}_{wt}^2)$, and note that

$$Var(\bar{\xi}_{wt}^2) = E(\bar{\xi}_{wt}^4) - \left[E(\bar{\xi}_{wt}^2)\right]^2.$$

$$E(\bar{\xi}_{wt}^4) = \sum_{i,j,k,l} w_i w_j w_k w_l E\left(\xi_{it} \xi_{jt} \xi_{kt} \xi_{lt}\right),$$

where Under assumptions 3 and 4

$$
\begin{aligned}
E\left(\xi_{it} \xi_{jt} \xi_{kt} \xi_{lt}\right) &= E\left(\varepsilon_{it}^4\right) + E\left[\left(\mathbf{x}_{it}' \boldsymbol{\upsilon}_i\right)^4\right] + 6\sigma_i^2 E(\mathbf{x}_{it}' \Omega \mathbf{x}_{it}), \text{ for } i = j = k = l \\
&= \left[\sigma_i^2 + E(\mathbf{x}_{it}' \Omega \mathbf{x}_{it})\right]\left[\sigma_k^2 + E(\mathbf{x}_{kt}' \Omega \mathbf{x}_{kt})\right], \text{ for } i = j, \text{ and } k = l \\
&= 0 \qquad \text{otherwise.}
\end{aligned}
$$

Hence

$$
\begin{aligned}
E(\bar{\xi}_{wt}^4) &= \sum_{i=1}^N w_i^4 \left\{E\left(\varepsilon_{it}^4\right) + E\left[\left(\mathbf{x}_{it}' \boldsymbol{\upsilon}_i\right)^4\right] + 6\sigma_i^2 E(\mathbf{x}_{it}' \Omega \mathbf{x}_{it})\right\} + \\
&\quad 3\left\{\sum_{i=1}^N w_i^2 \left[\sigma_i^2 + E(\mathbf{x}_{it}' \Omega \mathbf{x}_{it})\right]\right\}^2,
\end{aligned}
$$

and together with (A.7) yields

$$
\begin{aligned}
Var(\bar{\xi}_{wt}^2) &= \sum_{i=1}^N w_i^4 \left\{E\left(\varepsilon_{it}^4\right) + E\left(\mathbf{x}_{it}' \boldsymbol{\upsilon}_i\right)^4 + 6\sigma_i^2 E(\mathbf{x}_{it}' \Omega \mathbf{x}_{it})\right\} + \\
&\quad 2\left\{\sum_{i=1}^N w_i^2 \left[\sigma_i^2 + E(\mathbf{x}_{it}' \Omega \mathbf{x}_{it})\right]\right\}^2.
\end{aligned}
$$

However, by Cauchy-Schwarz inequality

$$\left(\mathbf{x}_{it}'\boldsymbol{\upsilon}_i\right)^2 \leq \left(\mathbf{x}_{it}'\mathbf{x}_{it}\right)\left(\boldsymbol{\upsilon}_i'\boldsymbol{\upsilon}_i\right),$$

and since $\mathbf{x}_{it}$ and $\boldsymbol{\upsilon}_i$ are independently distributed

$$E\left(\mathbf{x}_{it}'\boldsymbol{\upsilon}_i\right)^4 \leq E\left(\mathbf{x}_{it}'\mathbf{x}_{it}\right)^2 E\left(\boldsymbol{\upsilon}_i'\boldsymbol{\upsilon}_i\right)^2.$$

Also

$$E(\mathbf{x}_{it}'\Omega\mathbf{x}_{it}) \leq \lambda_{\max}(\Omega)E(\mathbf{x}_{it}'\mathbf{x}_{it})$$

Using these results and noting that under our assumptions, ε_{it}, $\mathbf{x}_{it}$, and $\boldsymbol{\upsilon}_i$ have finite fourth-order moments we have

$$Var(\bar{\xi}_{wt}^2) \leq K_1 \sum_{i=1}^{N} w_i^4 + K_2\left(\sum_{i=1}^{N} w_i^2\right)^2,$$

for some positive constants K_1 and K_2. Finally, noting that

$$\sum_{i=1}^{N} w_i^4 \leq \left(\sum_{i=1}^{N} w_i^2\right)^2 = O\left(\frac{1}{N^2}\right),$$

we have the desired result

$$Var(\bar{\xi}_{wt}^2) \leq (K_1 + K_2)\left(\sum_{i=1}^{N} w_i^2\right)^2 = O\left(\frac{1}{N^2}\right).$$

Consider now,

$$\begin{aligned} Var\left(\frac{1}{T}\sum_{t=1}^{T}\bar{\xi}_{wt}\right) &= \frac{1}{T^2}\sum_{t=1}^{T}\sum_{t'=1}^{T} E(\bar{\xi}_{wt}\bar{\xi}_{wt'}), \\ &= \frac{1}{T^2}\sum_{t=1}^{T}\sum_{t'=1}^{T}\sum_{i=1}^{N} w_i^2 E(\xi_{it}\xi_{it'}). \end{aligned}$$

But

$$\begin{aligned} E(\xi_{it}\xi_{it'}) &= \sigma_i^2 + E(\mathbf{x}_{it}'\Omega\mathbf{x}_{it}), \text{ for } t = t', \\ &= E(\mathbf{x}_{it}'\Omega\mathbf{x}_{it'}), \qquad \text{for } t \neq t'. \end{aligned}$$

Hence

$$Var\left(\frac{1}{T}\sum_{t=1}^{T}\bar{\xi}_{wt}\right) = \frac{1}{T}\sum_{i=1}^{N} w_i^2\sigma_i^2 + \frac{1}{T^2}\sum_{t=1}^{T}\sum_{t'=1}^{T}\sum_{i=1}^{N} w_i^2 E(\mathbf{x}_{it}'\Omega\mathbf{x}_{it'}).$$

The second term in the above expression can also be written as

$$\frac{1}{T^2}\sum_{t=1}^{T}\sum_{t'=1}^{T}\sum_{i=1}^{N} w_i^2 E(\mathbf{x}_{it}'\Omega\mathbf{x}_{it'}) = \sum_{i=1}^{N} w_i^2 E(\tilde{\mathbf{x}}_{iT}'\Omega\tilde{\mathbf{x}}_{iT}),$$

where

$$\tilde{\mathbf{x}}_{iT} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_{it},$$

is the time average of the regressors for the i^{th} unit. Therefore,

$$\begin{aligned} Var\left(\frac{1}{T}\sum_{t=1}^{T}\bar{\xi}_{wt}\right) &= \frac{1}{T}\sum_{i=1}^{N}w_i^2\sigma_i^2 + \sum_{i=1}^{N}w_i^2 E(\tilde{\mathbf{x}}_{iT}'\Omega\tilde{\mathbf{x}}_{iT}) \\ &\le K\left(\frac{\sum_{i=1}^{N}w_i^2}{T}\right) + \lambda_{\max}(\Omega)\sum_{i=1}^{N}w_i^2 E(\tilde{\mathbf{x}}_{iT}'\tilde{\mathbf{x}}_{iT}). \end{aligned}$$

But since for each i, $\mathbf{x}_{it}$ is covariance stationary with absolute summable autocovariances and zero means then $E(\tilde{\mathbf{x}}_{iT}'\tilde{\mathbf{x}}_{iT}) = O(\frac{1}{T})$.[17] Using this result and recalling that $\lambda_{\max}(\Omega)$ is bounded in N we have

$$Var\left(\frac{1}{T}\sum_{t=1}^{T}\bar{\xi}_{wt}\right) = O\left(\frac{\sum_{i=1}^{N}w_i^2}{T}\right) = O\left(\frac{1}{N\,T}\right).$$

It is now easily seen that $Cov\left(\bar{\xi}_{wt}\varepsilon_{it}, \bar{\xi}_{wt'}\varepsilon_{it'}\right) = 0$, for all $t \neq t'$,

$$Var\left(\frac{1}{T}\sum_{t=1}^{T}\bar{\xi}_{wt}\varepsilon_{it}\right) = \frac{1}{T^2}\sum_{t=1}^{T}Var\left(\bar{\xi}_{wt}\varepsilon_{it}\right),$$

which upon using (A.4) yields (A.8).

To prove (A.9) first note that by assumption $\bar{\xi}_{wt}^2$ is a covariance stationary process and $Var(\bar{\xi}_{wt}^2) = Var(\bar{\xi}_{wt'}^2)$ for all t and t'. Hence

$$Cov(\bar{\xi}_{wt}^2, \bar{\xi}_{wt'}^2) \le Var(\bar{\xi}_{wt}^2),$$

and

$$Var\left(\frac{1}{T}\sum_{t=1}^{T}\bar{\xi}_{wt}^2\right) = \frac{1}{T^2}\sum_{t=1}^{T}\sum_{t'=1}^{T}Cov(\bar{\xi}_{wt}^2, \bar{\xi}_{wt'}^2) \le Var(\bar{\xi}_{wt}^2),$$

and using (A.5) the desired result follows.[18]

Result (A.10) follows from (A.3).

To prove (A.11) first note that $\mathbf{x}_{it}$ and $\bar{\xi}_{wt}$ are independently distributed and $E\left(\mathbf{x}_{it}\bar{\xi}_{wt}\right) = \mathbf{0}$. Hence

$$Var\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_{it}\bar{\xi}_{wt}\right) = \frac{1}{T^2}\sum_{t=1}^{T}\sum_{t'=1}^{T}\sum_{k=1}^{N}\sum_{l=1}^{N}w_i w_k E\left(\mathbf{x}_{it}\mathbf{x}_{it'}'\right)E\left(\xi_{kt}\xi_{lt'}\right).$$

But

$$E\left(\xi_{kt}\xi_{lt'}\right) = 0 \text{ for } k \neq l, \text{ and all } t,$$

[17] See, for example, Proposition 10.5 in Hamilton (1994, p.279).

[18] In the case where the slope coefficients of the individual specific regressors are homogeneous, then $\bar{\boldsymbol{\xi}}_{wt}^2 = \bar{\varepsilon}_{wt}^2$ will be serially independent and $Var\left(\frac{1}{T}\sum_{t=1}^{T}\bar{\xi}_{wt}^2\right) = Var\left(\bar{\varepsilon}_{wt}^2\right)/T = O(1/N^2T)$.

and

$$
\begin{aligned}
E\left(\xi_{kt}\xi_{kt'}\right) &= \sigma_k^2 + E(\mathbf{x}'_{kt}\Omega\mathbf{x}_{kt'}), \text{ for } t = t' \\
&= E(\mathbf{x}'_{kt}\Omega\mathbf{x}_{kt'}) \text{ for } t \neq t'.
\end{aligned}
$$

Hence

$$
\begin{aligned}
Var\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_{it}\bar{\xi}_{wt}\right) &= \frac{1}{T}\sum_{k=1}^{N}w_k^2\sigma_k^2\left(\frac{\sum_{t=1}^{T}E\left(\mathbf{x}_{it}\mathbf{x}'_{it}\right)}{T}\right) + \\
&\quad \frac{1}{T}\sum_{k=1}^{N}w_k^2\left(\frac{\sum_{t=1}^{T}\sum_{t'=1}^{T}E\left(\mathbf{x}_{it}\mathbf{x}'_{it'}\right)E(\mathbf{x}'_{kt}\Omega\mathbf{x}_{kt'})}{T}\right).
\end{aligned}
$$

Since under our assumptions $\sigma_k^2 < \infty$, and $\mathbf{x}_{it}$ is covariance stationary then

$$
\frac{\sum_{t=1}^{T}E\left(\mathbf{x}_{it}\mathbf{x}'_{it}\right)}{T} = \Gamma_{x_i}(0) = O(1),
$$

and

$$
\frac{\sum_{t=1}^{T}\sum_{t'=1}^{T}E\left(\mathbf{x}_{it}\mathbf{x}'_{it'}\right)E(\mathbf{x}'_{kt}\Omega\mathbf{x}_{kt'})}{T} = \frac{\sum_{t=1}^{T}\sum_{t'=1}^{T}\Gamma_{x_i}(|t-t'|)Tr([\Omega\Gamma_{x_k}(|t-t'|)]}{T}, \tag{A.16}
$$

where

$$
E\left(\mathbf{x}_{it}\mathbf{x}'_{it'}\right) = \Gamma_{x_i}(\left|t-t'\right|).
$$

Furthermore, since by assumption $\mathbf{x}_{it}$ has absolute summable autocovariances, using standard results in the time series literature, it follows that (A.16) is bounded in T. Hence

$$
Var\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_{it}\bar{\xi}_{wt}\right) = O\left(\frac{1}{T}\sum_{k=1}^{N}w_k^2\right) = O\left(\frac{1}{N\,T}\right).
$$

A proof of (A.12) can be established along similar lines, making note of the following two properties. First, $\bar{\mathbf{g}}_{wt}$ and $\bar{\xi}_{wt}$ are independently distributed and $E(\bar{\mathbf{g}}_{wt}\bar{\xi}_{wt}) = 0$. Second since $\sum_{i=1}^{N}|w_i| < K$, and $(\mathbf{x}'_{it}, \mathbf{z}'_t, f_t)$, $i = 1, 2, ..., N$ are covariance stationary with absolute summable autocovariances, then $\bar{\mathbf{g}}_{wt} = (\mathbf{z}'_t, \bar{\mathbf{x}}'_{wt}, f_t)'$ will also be covariance stationary with absolute summable autocovariances for all N. ■

Proposition A.1 *Under assumption 1(a), 3, 4 and 5(a), for each t,*

$$
\bar{y}_{wt} - \left(\mathbf{f}'_t\bar{\boldsymbol{\alpha}}_w + \bar{\mathbf{x}}'_{wt}\boldsymbol{\beta} + \bar{\gamma}_w f_t\right) \overset{q.m.}{\to} 0, \text{ as } N \to \infty, \tag{A.17}
$$

where $\bar{\boldsymbol{\alpha}}_w = \sum_{j=1}^{N} w_j\boldsymbol{\alpha}_j$ *and* $\bar{\gamma}_w = \sum_{j=1}^{N} w_j\gamma_j$, *and* $\overset{q.m.}{\to}$ *denotes convergence in quadratic mean (or mean square error).*

Proof. Aggregating the individual-specific relations (2.1) and (2.2), using the weights, $\{w_j\}$, we obtain

$$\bar{y}_{wt} = \mathbf{z}_t'\bar{\boldsymbol{\alpha}}_w + \sum_{j=1}^{N} w_j \mathbf{x}_{jt}'\boldsymbol{\beta}_j + \bar{\gamma}_w f_t + \sum_{j=1}^{N} w_j \varepsilon_{jt}. \tag{A.18}$$

Assuming that the individual-specific slopes, $\boldsymbol{\beta}_j$, satisfy the random coefficient model, (2.5), we now have

$$\bar{y}_{wt} = \mathbf{z}_t'\bar{\alpha}_w + \bar{\mathbf{x}}_{wt}'\boldsymbol{\beta} + \bar{\gamma}_w f_t + \bar{\xi}_{wt}, \tag{A.19}$$

where $\bar{\xi}_{wt}$ is defined by (A.6). Therefore,

$$\lim_{N\to\infty} E\left[\bar{y}_{wt} - \left(\mathbf{z}_t'\bar{\alpha}_w + \bar{\mathbf{x}}_{wt}'\boldsymbol{\beta} + \bar{\gamma}_w f_t\right)\right]^2 = \lim_{N\to\infty} E\left(\bar{\xi}_{wt}^2\right).$$

However, using (A.15) we have

$$\lim_{N\to\infty} E\left(\bar{\xi}_{wt}^2\right) \leq K \lim_{N\to\infty}\left(\sum_{j=1}^{N} w_j^2\right) = 0,$$

which establishes that

$$\bar{y}_{wt} - \left(\mathbf{z}_t'\bar{\alpha}_w + \bar{\mathbf{x}}_{wt}'\boldsymbol{\beta} + \bar{\gamma}_w f_t\right) \overset{q.m.}{\to} 0, \text{ as } N \to \infty.$$

■

Lemma A.2 *Under assumptions 1(a), (3), (4), and 5(a) we have*

$$\frac{\bar{\boldsymbol{\xi}}_w'\bar{\boldsymbol{\xi}}_w}{\sqrt{T}} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\bar{\xi}_{wt}^2 = O_p\left(\frac{\sqrt{T}}{N}\right), \tag{A.20}$$

$$\frac{\boldsymbol{\varepsilon}_i'\bar{\boldsymbol{\xi}}_w}{\sqrt{T}} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{it}\bar{\xi}_{wt} = O_p\left(\frac{\sqrt{T}}{N}\right), \tag{A.21}$$

$$\frac{\mathbf{X}_i'\bar{\boldsymbol{\xi}}_w}{T} = \frac{\sum_{t=1}^{T}\mathbf{x}_{it}\bar{\xi}_{wt}}{T} = O_p\left(\frac{1}{\sqrt{NT}}\right) \tag{A.22}$$

$$\frac{\bar{\mathbf{G}}_w'\bar{\boldsymbol{\xi}}_w}{T} = \frac{\sum_{t=1}^{T}\bar{\mathbf{g}}_{wt}\bar{\xi}_{wt}}{T} = O_p\left(\frac{1}{\sqrt{NT}}\right) \tag{A.23}$$

$$\frac{\bar{\mathbf{H}}_w'\bar{\mathbf{H}}_w}{T} = \mathbf{A}_w\left(\frac{\mathbf{G}_\mu'\mathbf{G}_\mu}{T}\right)\mathbf{A}_w' + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right), \tag{A.24}$$

$$\frac{\bar{\mathbf{H}}_w'\bar{\mathbf{G}}_w}{T} = \mathbf{A}_w\left(\frac{\mathbf{G}_\mu'\mathbf{G}_\mu}{T}\right) + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right), \tag{A.25}$$

$$\frac{\bar{\mathbf{H}}_w'\mathbf{X}_i}{T} = \mathbf{A}_w\left(\frac{\mathbf{G}_\mu'\mathbf{X}_i}{T}\right) + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right), \tag{A.26}$$

$$\frac{\bar{\mathbf{H}}_w'\bar{\boldsymbol{\xi}}_w}{T} = O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right), \tag{A.27}$$

$$\frac{\bar{\mathbf{H}}_w'\boldsymbol{\varepsilon}_i}{T} = \mathbf{A}_w\left(\frac{\mathbf{G}_\mu'\boldsymbol{\varepsilon}_i}{T}\right) + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{N\sqrt{T}}\right), \tag{A.28}$$

and

$$\frac{\mathbf{G}_\mu'\boldsymbol{\varepsilon}_i}{T} = O_p\left(\frac{1}{\sqrt{T}}\right). \tag{A.29}$$

where $\bar{\mathbf{H}}_w$, $\mathbf{A}_w$, *and* $\bar{\xi}_{wt}$ *are defined by (4.14), (4.23), and (A.6), respectively,* $\bar{\mathbf{g}}_{wt} = (\mathbf{z}_t', \bar{\mathbf{x}}_{wt}', f_t)'$, $\mathbf{g}_{\mu t} = (\mathbf{z}_t', \boldsymbol{\mu}_{tx}', f_t)'$, *and* $\mathbf{G}_\mu' = (\mathbf{g}_{\mu 1}, \mathbf{g}_{\mu 2}, ..., \mathbf{g}_{\mu T})$, *where* $\boldsymbol{\mu}_{tx} = p\lim_{N\to\infty}(\bar{\mathbf{x}}_{wt})$.[19]

Proof. Using (A.9) and (A.10)

$$Var\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\bar{\xi}_{wt}^2\right) \leq O\left(\frac{T}{N^2}\right), \text{ and } E\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\bar{\xi}_{wt}^2\right) = O\left(\sqrt{T}\sum_{j=1}^{N}w_j^2\right),$$

and (A.20) follows noting that $\sum_{j=1}^{N} w_j^2 = O(1/N)$.

Similarly, using (A.2) and (A.8) we have

$$Var\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\bar{\xi}_{wt}\varepsilon_{it}\right) = O\left(w_i^2\right), \ E\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\bar{\xi}_{wt}\varepsilon_{it}\right) = O\left(\sqrt{T}w_i\right),$$

and since by assumption $w_i = O(1/N)$, then (A.21) follows. Results (A.22) and (A.23) can also be readily established using (A.11) and (A.12) and noting that $\bar{\mathbf{g}}_{wt}$ and $\bar{\xi}_{wt}$ are independently distributed.

To prove (A.24), using (4.22) and (4.23) we first note that

$$\begin{aligned}\frac{\bar{\mathbf{H}}_w'\bar{\mathbf{H}}_w}{T} &= \sum_{t=1}^{T}\bar{\mathbf{h}}_{wt}\bar{\mathbf{h}}_{wt}' = \mathbf{A}_w\left(\frac{\sum_{t=1}^{T}\bar{\mathbf{g}}_{wt}\bar{\mathbf{g}}_{wt}'}{T}\right)\mathbf{A}_w' + \mathbf{A}_w\left(\frac{\sum_{t=1}^{T}\bar{\mathbf{g}}_{wt}\bar{\boldsymbol{\nu}}_{wt}'}{T}\right) \\ &\quad \left(\frac{\sum_{t=1}^{T}\bar{\boldsymbol{\nu}}_{wt}\bar{\mathbf{g}}_{wt}'}{T}\right)\mathbf{A}_w' + \left(\frac{\sum_{t=1}^{T}\bar{\boldsymbol{\nu}}_{wt}\bar{\boldsymbol{\nu}}_{wt}'}{T}\right),\end{aligned} \tag{A.30}$$

[19]For expositional simplicity here we are assuming that none of the individual specific regressors converge in probability to non-stochastic constants under cross section averaging.

where under assumption 5(a) $\mathbf{A}_w$ is a non-singular matrix. Also

$$\frac{\sum_{t=1}^{T}\bar{\mathbf{g}}_{wt}\bar{\boldsymbol{\nu}}'_{wt}}{T} = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \frac{\sum_{t=1}^{T}\mathbf{z}_t\bar{\xi}_{wt}}{T} \\ \mathbf{0} & \mathbf{0} & \frac{\sum_{t=1}^{T}\bar{\mathbf{x}}_{wt}\bar{\xi}_{wt}}{T} \\ \mathbf{0} & \mathbf{0} & \frac{\sum_{t=1}^{T}f_t\bar{\xi}_{wt}}{T} \end{pmatrix} = \begin{pmatrix} \mathbf{0} & \frac{\sum_{t=1}^{T}\bar{\mathbf{g}}_{wt}\bar{\xi}_{wt}}{T} \end{pmatrix}, \tag{A.31}$$

and

$$\frac{\sum_{t=1}^{T}\bar{\boldsymbol{\nu}}_{wt}\bar{\boldsymbol{\nu}}'_{wt}}{T} = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \frac{\sum_{t=1}^{T}\bar{\xi}_{wt}^2}{T} \end{pmatrix}. \tag{A.32}$$

Using (A.20) and (A.23) we now have

$$\frac{\sum_{t=1}^{T}\bar{\boldsymbol{\nu}}_{wt}\bar{\boldsymbol{\nu}}'_{wt}}{T} = O_p\left(\frac{1}{N}\right) \text{, for all } T, \tag{A.33}$$

and

$$\frac{\sum_{t=1}^{T}\bar{\mathbf{g}}_{wt}\bar{\boldsymbol{\nu}}'_{wt}}{T} = O_p\left(\frac{1}{\sqrt{NT}}\right). \tag{A.34}$$

Using these results in (A.30) yields (note that $\mathbf{A}_w$ is a fixed matrix):

$$\frac{\bar{\mathbf{H}}'_w\bar{\mathbf{H}}_w}{T} = \mathbf{A}_w\left(\frac{\sum_{t=1}^{T}\bar{\mathbf{g}}_{wt}\bar{\mathbf{g}}'_{wt}}{T}\right)\mathbf{A}'_w + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right). \tag{A.35}$$

But under assumption 5(b)[20]

$$\bar{\mathbf{g}}_{wt} = \mathbf{g}_{\mu t} + O_p\left(\frac{1}{\sqrt{N}}\right) \tag{A.36}$$

then (for all T)

$$\frac{\sum_{t=1}^{T}\bar{\mathbf{g}}_{wt}\bar{\mathbf{g}}'_{wt}}{T} = \frac{\sum_{t=1}^{T}\mathbf{g}_{\mu t}\mathbf{g}'_{\mu t}}{T} + O_p\left(\frac{1}{\sqrt{N}}\right), \tag{A.37}$$

and

$$\frac{\sum_{t=1}^{T}\bar{\mathbf{g}}_{wt}\mathbf{x}'_{it}}{T} = \frac{\sum_{t=1}^{T}\mathbf{g}_{\mu t}\mathbf{x}'_{it}}{T} + O_p\left(\frac{1}{\sqrt{N}}\right). \tag{A.38}$$

Hence

$$\frac{\bar{\mathbf{H}}'_w\bar{\mathbf{H}}_w}{T} = \mathbf{A}_w\left(\frac{\sum_{t=1}^{T}\mathbf{g}_{\mu t}\mathbf{g}'_{\mu t}}{T}\right)\mathbf{A}'_w + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right). \tag{A.39}$$

[20]For expositional simplicity here we are assuming that none of the individual specific regressors converge in probability to non-stochastic constants under cross section averaging.

which establishes (A.24). (A.25) also follows along similar lines.

Consider now

$$\frac{\bar{\mathbf{H}}_w'\mathbf{X}_i}{T} = \frac{\sum_{t=1}^T \bar{\mathbf{h}}_{wt}\mathbf{x}_{it}'}{T} = \mathbf{A}_w\left(\frac{\sum_{t=1}^T \bar{\mathbf{g}}_{wt}\mathbf{x}_{it}'}{T}\right) + \left(\frac{\sum_{t=1}^T \bar{\boldsymbol{\nu}}_{wt}\mathbf{x}_{it}'}{T}\right), \tag{A.40}$$

and using (A.8), (A.11) and (A.38) we have

$$\begin{aligned}\frac{\bar{\mathbf{H}}_w'\mathbf{X}_i}{T} &= \mathbf{A}_w\left(\frac{\sum_{t=1}^T \bar{\mathbf{g}}_{wt}\mathbf{x}_{it}'}{T}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right)\\ &= \mathbf{A}_w\left(\frac{\sum_{t=1}^T \mathbf{g}_{\mu t}\mathbf{x}_{it}'}{T}\right) + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right).\end{aligned}$$

Similarly,

$$\frac{\bar{\mathbf{H}}_w'\bar{\boldsymbol{\xi}}_w}{T} = \left(\frac{\sum_{t=1}^T \bar{\mathbf{h}}_{wt}\bar{\xi}_{wt}}{T}\right) = \mathbf{A}_w\left(\frac{\sum_{t=1}^T \bar{\mathbf{g}}_{wt}\bar{\xi}_{wt}}{T}\right) + \left(\frac{\sum_{t=1}^T \bar{\boldsymbol{\nu}}_{wt}\bar{\xi}_{wt}}{T}\right), \tag{A.41}$$

where

$$\frac{\sum_{t=1}^T \bar{\boldsymbol{\nu}}_{wt}\bar{\xi}_{wt}}{T} = \begin{pmatrix}\mathbf{0}\\ \mathbf{0}\\ \frac{\sum_{t=1}^T \bar{\xi}_{wt}^2}{T}\end{pmatrix} = O_p\left(\frac{1}{N}\right).$$

Hence using (A.34) we have

$$\frac{\bar{\mathbf{H}}_w'\bar{\boldsymbol{\xi}}_w}{T} = O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{N}\right). \tag{A.42}$$

Finally,

$$\frac{\bar{\mathbf{H}}_w'\boldsymbol{\varepsilon}_i}{T} = \frac{\sum_{t=1}^T \bar{\mathbf{h}}_{wt}\varepsilon_{it}}{T} = \mathbf{A}_w\left(\frac{\sum_{t=1}^T \bar{\mathbf{g}}_{wt}\varepsilon_{it}}{T}\right) + \left(\frac{\sum_{t=1}^T \bar{\boldsymbol{\nu}}_{wt}\varepsilon_{it}}{T}\right), \tag{A.43}$$

where (using (A.8))

$$\frac{\sum_{t=1}^T \bar{\boldsymbol{\nu}}_{wt}\varepsilon_{it}}{T} = \begin{pmatrix}\mathbf{0}\\ \mathbf{0}\\ \frac{\sum_{t=1}^T \bar{\xi}_{wt}\varepsilon_{it}}{T}\end{pmatrix} = O_p\left(\frac{1}{N\sqrt{T}}\right), \tag{A.44}$$

and

$$\frac{\sum_{t=1}^T \bar{\mathbf{g}}_{wt}\varepsilon_{it}}{T} = \frac{\sum_{t=1}^T \mathbf{g}_{\mu t}\varepsilon_{it}}{T} + O_p\left(\frac{1}{\sqrt{N}}\right),$$

hence

$$\frac{\bar{\mathbf{H}}_w'\boldsymbol{\varepsilon}_i}{T} = \mathbf{A}_w\left(\frac{\sum_{t=1}^T \mathbf{g}_{\mu t}\varepsilon_{it}}{T}\right) + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{N\sqrt{T}}\right).$$

Finally, noting that under assumption 5(b) $T^{-1}\sum_{t=1}^{T} E\left(\mathbf{g}_{\mu t}\mathbf{g}'_{\mu t}\right)$ exists and is bounded in N and T, we have

$$\frac{\mathbf{G}'_{\mu}\boldsymbol{\varepsilon}_i}{T} = \frac{\sum_{t=1}^{T}\mathbf{g}_{\mu t}\varepsilon_{it}}{T} = O_p\left(\frac{1}{\sqrt{T}}\right).$$

■

Proof of Theorem 4.1

Part (a) - N**-asymptotic**. In this part T is fixed and the limits are taken with $N \to \infty$. Using (A.24) and (A.26) and noting that under assumption 5(a) $\mathbf{A}_w$ is a non-singular matrix we have

$$\mathbf{X}'_i\bar{\mathbf{M}}_w\mathbf{X}_i = \mathbf{X}'_i\mathbf{X}_i - \mathbf{X}'_i\bar{\mathbf{H}}_w\left(\bar{\mathbf{H}}'_w\bar{\mathbf{H}}_w\right)^{-1}\bar{\mathbf{H}}'_w\mathbf{X}_i = \mathbf{X}'_i\mathbf{M}_{\mu}\mathbf{X}_i + O_p\left(\frac{1}{\sqrt{N}}\right), \tag{A.45}$$

where

$$\mathbf{M}_{\mu} = \mathbf{I}_T - \mathbf{G}_{\mu}(\mathbf{G}'_{\mu}\mathbf{G}_{\mu})^{-1}\mathbf{G}'_{\mu}. \tag{A.46}$$

Similarly, using results in Lemma A.2 it is easily seen that

$$\mathbf{X}'_i\bar{\mathbf{M}}_w\bar{\boldsymbol{\xi}}_w \xrightarrow{p} 0, \tag{A.47}$$

and

$$\mathbf{X}'_i\bar{\mathbf{M}}_w\boldsymbol{\varepsilon}_i = \mathbf{X}'_i\mathbf{M}_{\mu}\boldsymbol{\varepsilon}_i + O_p\left(\frac{1}{\sqrt{N}}\right). \tag{A.48}$$

Substituting these results in (4.21) we now have[21]

$$\hat{\mathbf{b}}_i - \boldsymbol{\beta}_i = \left(\mathbf{X}'_i\mathbf{M}_{\mu}\mathbf{X}_i\right)^{-1}\mathbf{X}'_i\mathbf{M}_{\mu}\boldsymbol{\varepsilon}_i + O_p\left(\frac{1}{\sqrt{N}}\right), \tag{A.49}$$

where ε_{it} is distributed independently of $\mathbf{x}_{it}$ and $\mathbf{g}_{\mu t}$. Therefore,

$$\lim_{N\to\infty} E\left(\hat{\mathbf{b}}_i\right) = \boldsymbol{\beta}_i.$$

Also under the normality assumption for a fixed $T > k_z + 2k_x + 1$, and as $N \to \infty$

$$\hat{\mathbf{b}}_i - \boldsymbol{\beta}_i \xrightarrow{d} N(\mathbf{0}, \Sigma_{T,b_i}), \tag{A.50}$$

where

$$\Sigma_{T,b_i} = \sigma_i^2\left(\mathbf{X}'_i\mathbf{M}_{\mu}\mathbf{X}_i\right)^{-1}. \tag{A.51}$$

[21] Note that under assumption 5(b), $\mathbf{X}'_i\mathbf{M}_{\mu}\mathbf{X}_i$ is a positive definite matrix. A necessary order condition for this is given by $T > k_z + 2k_x + 1$.

Part (b) - **Joint N and T asymptotic**. In this part all probability limits are taken under $(N,T) \xrightarrow{j} \infty$. First note that by assumption 5(b) we have

$$\begin{aligned} \frac{\sum_{t=1}^{T} \mathbf{g}_{\mu t}\mathbf{g}_{\mu t}'}{T} &= \Sigma_{\mu\mu} + O_p\left(\frac{1}{\sqrt{T}}\right), \\ \frac{\sum_{t=1}^{T} \mathbf{g}_{\mu t}\mathbf{x}_{it}'}{T} &= \Sigma_{\mu x_i} + O_p\left(\frac{1}{\sqrt{T}}\right), \end{aligned} \tag{A.52}$$

and

$$\frac{\mathbf{X}_i'\mathbf{M}_\mu \mathbf{X}_i}{T} = \Psi_i + O_p\left(\frac{1}{\sqrt{T}}\right),$$

where

$$\Psi_i = \Sigma_{x_i x_i} - \Sigma_{x_i\mu}\Sigma_{\mu\mu}^{-1}\Sigma_{\mu x_i}, \tag{A.53}$$

is a non-stochastic positive definite matrix. See, in particular, (2.12). These in conjunction with the results established in Lemma A.2 now yield

$$\frac{\bar{\mathbf{H}}_w'\bar{\mathbf{H}}_w}{T} = \mathbf{A}_w\Sigma_{\mu\mu}\mathbf{A}_w' + O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{1}{\sqrt{N}}\right), \tag{A.54}$$

$$\frac{\bar{\mathbf{H}}_w'\mathbf{X}_i}{T} = \mathbf{A}_w\Sigma_{\mu x_i} + O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{1}{\sqrt{N}}\right), \tag{A.55}$$

$$\frac{\bar{\mathbf{H}}_w'\bar{\boldsymbol{\xi}}_w}{T} = O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right), \tag{A.56}$$

and noting that $\mathbf{A}_w$ is a non-singular matrix then

$$\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{X}_i}{T} \xrightarrow{p} \Psi_i, \tag{A.57}$$

and using (4.21) we have

$$\hat{\mathbf{b}}_i - \boldsymbol{\beta}_i \xrightarrow{p} -\left(\frac{\gamma_i}{\bar{\gamma}_w}\right)\Psi_i^{-1}\left(\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\bar{\boldsymbol{\xi}}_w}{T}\right) + \Psi_i^{-1}\left(\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\boldsymbol{\varepsilon}_i}{T}\right). \tag{A.58}$$

But

$$\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\bar{\boldsymbol{\xi}}_w}{T} = \frac{\mathbf{X}_i'\bar{\boldsymbol{\xi}}_w}{T} - \left(\frac{\mathbf{X}_i'\bar{\mathbf{H}}_w}{T}\right)\left(\frac{\bar{\mathbf{H}}_w'\bar{\mathbf{H}}_w}{T}\right)^{-1}\left(\frac{\bar{\mathbf{H}}_w'\bar{\boldsymbol{\xi}}_w}{T}\right),$$

and using (A.54) to (A.56), and (A.22) we have

$$\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\bar{\boldsymbol{\xi}}_w}{T} \xrightarrow{p} \mathbf{0}. \tag{A.59}$$

Also from (A.28) and (A.29)

$$\frac{\bar{\mathbf{H}}_w'\boldsymbol{\varepsilon}_i}{T} = O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{1}{\sqrt{N}}\right).$$

and since $\mathbf{X}_i'\varepsilon_i/T = O_p\left(T^{-1/2}\right)$ then

$$\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\varepsilon_i}{T} \xrightarrow{p} \mathbf{0}. \tag{A.60}$$

Using (A.59) and (A.60) in (A.58) the desired result follows, namely $\hat{\mathbf{b}}_i \xrightarrow{p} \boldsymbol{\beta}_i$, as $(N,T) \xrightarrow{j} \infty$ with no particular restrictions on the order by which N and/or T are allowed to tend to infinity.

To establish the asymptotic normality of $\hat{\mathbf{b}}_i$ stronger convergence results are required. First we need to show that the limiting distribution of $\sqrt{T}\left(\hat{\mathbf{b}}_i - \boldsymbol{\beta}_i\right)$ does not depend on the factor loadings, γ_i. As we shall see this is valid only if $\sqrt{T}/N \to 0$, as $(N,T) \to \infty$. To see this using (4.21) and (A.57) we have

$$\sqrt{T}\left(\hat{\mathbf{b}}_i - \boldsymbol{\beta}_i\right) \overset{d}{\sim} -\left(\frac{\gamma_i}{\bar{\gamma}_w}\right)\Psi_i^{-1}\left(\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\bar{\boldsymbol{\xi}}_w}{\sqrt{T}}\right) + \Psi_i^{-1}\left(\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\varepsilon_i}{\sqrt{T}}\right).$$

To ensure that the asymptotic distribution of $\sqrt{T}\left(\hat{\mathbf{b}}_i - \boldsymbol{\beta}_i\right)$ does not depend on the factor it is necessary that $T^{-1/2}\left(\mathbf{X}_i'\bar{\mathbf{M}}_w\bar{\boldsymbol{\xi}}_w\right) \xrightarrow{p} \mathbf{0}$ as $(N,T) \xrightarrow{j} \infty$.

$$\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\bar{\boldsymbol{\xi}}_w}{\sqrt{T}} = \left(\frac{\mathbf{X}_i'\bar{\boldsymbol{\xi}}_w}{\sqrt{T}}\right) - \left(\frac{\mathbf{X}_i'\bar{\mathbf{H}}_w}{T}\right)\left(\frac{\bar{\mathbf{H}}_w'\bar{\mathbf{H}}_w}{T}\right)^{-1}\left(\frac{\bar{\mathbf{H}}_w'\bar{\boldsymbol{\xi}}_w}{\sqrt{T}}\right). \tag{A.61}$$

From (A.22)

$$\left(\frac{\mathbf{X}_i'\bar{\boldsymbol{\xi}}_w}{\sqrt{T}}\right) = O_p\left(\frac{1}{\sqrt{N}}\right),$$

and using (A.24) and (A.26) we first observe that

$$\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\bar{\boldsymbol{\xi}}_w}{\sqrt{T}} \xrightarrow{p} -\Sigma_{x_i\mu}\mathbf{A}_w'\left(\mathbf{A}_w\Sigma_{\mu\mu}\mathbf{A}_w'\right)^{-1}\left(\frac{\bar{\mathbf{H}}_w'\bar{\boldsymbol{\xi}}_w}{\sqrt{T}}\right). \tag{A.62}$$

Hence, the rate of convergence of $T^{-1/2}\left(\mathbf{X}_i'\bar{\mathbf{M}}_w\bar{\boldsymbol{\xi}}_w\right)$ is determined by that of $T^{-1/2}\left(\bar{\mathbf{H}}_w'\bar{\boldsymbol{\xi}}_w\right)$. But from (A.27)

$$\left(\frac{\bar{\mathbf{H}}_w'\bar{\boldsymbol{\xi}}_w}{\sqrt{T}}\right) = O_p\left(\frac{\sqrt{T}}{N}\right) + O_p\left(\frac{1}{\sqrt{N}}\right),$$

Therefore, for the asymptotic distribution of distribution of $\sqrt{T}\left(\hat{\mathbf{b}}_i - \boldsymbol{\beta}_i\right)$to be free of nuisance parameters it is sufficient that $\sqrt{T}/N \to 0$, as $(N,T) \xrightarrow{j} \infty$. Under this additional condition

$$\sqrt{T}\left(\hat{\mathbf{b}}_i - \boldsymbol{\beta}_i\right) \overset{d}{\sim} \Psi_i^{-1}\left(\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\varepsilon_i}{\sqrt{T}}\right). \tag{A.63}$$

Again using (A.54), (A.55) and (A.28) as well as (A.43) and (4.10) we have

$$\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\varepsilon_i}{\sqrt{T}} \overset{d}{\sim} \frac{\mathbf{X}_i'\varepsilon_i}{\sqrt{T}} - \Sigma_{x_i\mu}\Sigma_{\mu\mu}^{-1}\left(\frac{\mathbf{G}_\mu'\varepsilon_i}{\sqrt{T}}\right), \tag{A.64}$$

and by application of standard central limit theorem for regression models it is easily seen that with $T \to \infty$,

$$\frac{1}{\sqrt{T}}\left(\mathbf{X}_i'\varepsilon_i - \Sigma_{x_i\mu}\Sigma_{\mu\mu}^{-1}\mathbf{G}_\mu'\varepsilon_i\right) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\mathbf{x}_{it} - \Sigma_{x_i\mu}\Sigma_{\mu\mu}^{-1}\mathbf{g}_{\mu t}\right)\varepsilon_{it} \xrightarrow{d} N(\mathbf{0},\sigma_i^2\Psi_i), \tag{A.65}$$

where Ψ_i is given by (A.57). Using this result in (A.63) now yields

$$\sqrt{T}\left(\hat{\mathbf{b}}_i - \boldsymbol{\beta}_i\right) \xrightarrow{d} N(\mathbf{0},\sigma_i^2\Psi_i^{-1}), \text{ as } (N,T) \xrightarrow{j} \infty \text{ and } \sqrt{T}/N \to 0. \tag{A.66}$$

as required.

Proof of Asymptotic Unbiasedness of $\hat{\Sigma}_{T,b_i}$

In view of (A.45) it is sufficient to show that $\lim_{N\to\infty} E(\hat{\sigma}_i^2) = \sigma_i^2$ for a fixed $T > k_z + 2k_x + 1$. Using (4.12), $\hat{\sigma}_i^2$ given by (4.31) can be written as

$$\hat{\sigma}_i^2 = \frac{\mathbf{y}_i'\bar{\mathbf{D}}_w\mathbf{y}_i}{T-(k_z+2k_x+1)}, \tag{A.67}$$

where

$$\bar{\mathbf{D}}_w = \bar{\mathbf{M}}_w - \bar{\mathbf{M}}_w\mathbf{X}_i\left(\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{X}_i\right)^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_w. \tag{A.68}$$

Under (4.15) and using (4.20)

$$\mathbf{y}_i'\bar{\mathbf{D}}_w\mathbf{y}_i = \left(\varepsilon_i - \frac{\gamma_i}{\bar{\gamma}_w}\bar{\boldsymbol{\xi}}_w\right)'\bar{\mathbf{D}}_w\left(\varepsilon_i - \frac{\gamma_i}{\bar{\gamma}_w}\bar{\boldsymbol{\xi}}_w\right). \tag{A.69}$$

But

$$\bar{\boldsymbol{\xi}}_w'\bar{\mathbf{D}}_w\varepsilon_i = \bar{\boldsymbol{\xi}}_w'\bar{\mathbf{M}}_w\varepsilon_i - \bar{\boldsymbol{\xi}}_w'\bar{\mathbf{M}}_w\mathbf{X}_i\left(\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{X}_i\right)^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_w\varepsilon_i, \tag{A.70}$$

and

$$\bar{\boldsymbol{\xi}}_w'\bar{\mathbf{D}}_w\bar{\boldsymbol{\xi}}_w = \bar{\boldsymbol{\xi}}_w'\bar{\mathbf{M}}_w\bar{\boldsymbol{\xi}}_w - \bar{\boldsymbol{\xi}}_w'\bar{\mathbf{M}}_w\mathbf{X}_i\left(\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{X}_i\right)^{-1}\mathbf{X}_i'\bar{\mathbf{M}}_w\bar{\boldsymbol{\xi}}_w'. \tag{A.71}$$

In the case where T is fixed and $N \to \infty$, using the various results in Lemma A.2 it is easily seen that

$$\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{X}_i \xrightarrow{p} \mathbf{X}_i'\mathbf{M}_\mu\mathbf{X}_i,\ \bar{\boldsymbol{\xi}}_w'\bar{\mathbf{D}}_w\varepsilon_i \xrightarrow{p} 0,\ \bar{\boldsymbol{\xi}}_w'\bar{\mathbf{D}}_w\bar{\boldsymbol{\xi}}_w \xrightarrow{p} 0,$$

and

$$\mathbf{y}_i'\bar{\mathbf{D}}_w\mathbf{y}_i \xrightarrow{p} \varepsilon_i'\mathbf{D}_\mu\varepsilon_i, \text{ as } N \to \infty, \text{ for a fixed } T,$$

where

$$\mathbf{D}_\mu = \mathbf{M}_\mu - \mathbf{M}_\mu\mathbf{X}_i\left(\mathbf{X}_i'\mathbf{M}_\mu\mathbf{X}_i\right)^{-1}\mathbf{X}_i'\mathbf{M}_\mu, \tag{A.72}$$

with all elements of $\boldsymbol{\varepsilon}_i$ being distributed independently of those of $\mathbf{D}_\mu$. Hence[22]

$$\lim_{N\to\infty} E(\hat{\sigma}_i^2) = \frac{E\left(\boldsymbol{\varepsilon}_i'\mathbf{D}_\mu\boldsymbol{\varepsilon}_i\right)}{T-(k_z+2k_x+1)} = \frac{\sigma_i^2 Tr(\mathbf{D}_\mu)}{T-(k_z+2k_x+1)} = \sigma_i^2.$$

Proof of Consistency of $\hat{\Sigma}_{b_i}$

In view of (A.57) it is sufficient to prove that $\hat{\sigma}_i^2 \xrightarrow{p} \sigma_i^2$, as $(N,T) \xrightarrow{j} \infty$. First using (A.70) and (A.71) and the results in Lemma A.2 we have

$$\frac{\bar{\boldsymbol{\xi}}_w'\bar{\mathbf{D}}_w\bar{\boldsymbol{\xi}}_w}{T} \xrightarrow{p} 0, \text{ and } \frac{\bar{\boldsymbol{\xi}}_w'\bar{\mathbf{D}}_w\boldsymbol{\varepsilon}_i}{T} \xrightarrow{p} 0, \text{ as } (N,T) \xrightarrow{j} \infty.$$

Hence

$$\hat{\sigma}_i^2 = \frac{\mathbf{y}_i'\bar{\mathbf{D}}_w\mathbf{y}_i}{T-(k_z+2k_x+1)} \xrightarrow{p} \frac{\boldsymbol{\varepsilon}_i'\bar{\mathbf{D}}_w\boldsymbol{\varepsilon}_i}{T-(k_z+2k_x+1)} \xrightarrow{p} \frac{\boldsymbol{\varepsilon}_i'\mathbf{D}_\mu\boldsymbol{\varepsilon}_i}{T-(k_z+2k_x+1)},$$

where $\mathbf{D}_\mu$ is defined by (A.72) and we have the desired result, $\hat{\sigma}_i^2 \xrightarrow{p} \sigma_i^2$, as $(N,T) \xrightarrow{j} \infty$.

Proof of Theorem 4.2

Part (a) - N**-asymptotic with** T **fixed**. Throughout note that by assumption $T > 2k_x+k_z+1$. Let

$$\Delta_N = \bar{\mathbf{M}}_w - \mathbf{M}_\mu, \tag{A.73}$$

and note that

$$\left\|\sum_{i=1}^N \theta_i\mathbf{X}_i'\Delta_N\mathbf{X}_i\right\| \leq \left(\sum_{i=1}^N |\theta_i| \left\|\mathbf{X}_i\right\|^2\right) \left\|\Delta_N\right\|. \tag{A.74}$$

But $\left\|\mathbf{X}_i\right\|^2 = Tr\left(\mathbf{X}_i'\mathbf{X}_i\right) = \sum_{t=1}^T \mathbf{x}_{it}'\mathbf{x}_{it}$ and under assumption 1(a) and recalling from (**??**) that $\sum_{i=1}^N |\theta_i| < \infty$, we have

$$E\left(\sum_{i=1}^N |\theta_i| \left\|\mathbf{X}_i\right\|^2\right) = \sum_{t=1}^T\sum_{i=1}^N |\theta_i| E\left(\mathbf{x}_{it}'\mathbf{x}_{it}\right) \leq K\sum_{i=1}^N |\theta_i| < \infty. \tag{A.75}$$

Also

$$\Delta_N = \mathbf{G}_\mu(\mathbf{G}_\mu'\mathbf{G}_\mu)^{-1}\mathbf{G}_\mu' - \bar{\mathbf{H}}_w\left(\bar{\mathbf{H}}_w'\bar{\mathbf{H}}_w\right)^{-1}\bar{\mathbf{H}}_w',$$

and noting that $Tr\left(\Delta_N\right) = 0$, it is easily seen that

$$\left\|\Delta_N\right\|^2 = 2\; Tr\left[\left(\frac{\mathbf{G}_\mu'\mathbf{G}_\mu}{T}\right)^{-1}\left(\frac{\mathbf{G}_\mu'\bar{\mathbf{M}}_w\mathbf{G}_\mu}{T}\right)\right].$$

Using (A.24) and (A.25) and under assumption 5(b) it is now easily seen that (for all T)

$$\frac{\mathbf{G}_\mu'\bar{\mathbf{M}}_w\mathbf{G}_\mu}{T} = O_p\left(\frac{1}{\sqrt{N}}\right), \tag{A.76}$$

[22]Note that $Tr(\mathbf{M}_\mu) = T-(k_z+k_x+1)$, and $Tr\left(\mathbf{M}_\mu\mathbf{X}_i\left(\mathbf{X}_i'\mathbf{M}_\mu\mathbf{X}_i\right)^{-1}\mathbf{X}_i'\mathbf{M}_\mu\right) = k_x$.

and hence

$$\|\Delta_N\|^2 = O_p\left(\frac{1}{\sqrt{N}}\right). \tag{A.77}$$

Therefore, using (A.75) and (A.77), from (A.74) it follows that for a fixed T and as $N \to \infty$

$$\sum_{i=1}^{N} \theta_i \mathbf{X}_i' \Delta_N \mathbf{X}_i \xrightarrow{p} \mathbf{0}, \tag{A.78}$$

and under assumption 5(c)

$$\sum_{i=1}^{N} \theta_i \mathbf{X}_i' \bar{\mathbf{M}}_w \mathbf{X}_i \ \xrightarrow{p}\ \sum_{i=1}^{N} \theta_i \mathbf{X}_i' \mathbf{M}_\mu \mathbf{X}_i \xrightarrow{p} \Psi_T, \text{ as } N \to \infty, \tag{A.79}$$

where for a fixed T, Ψ_T is a non-stochastic positive definite matrix.

Using this result in (4.40)

$$\hat{\mathbf{b}} - \boldsymbol{\beta} \ \xrightarrow{p}\ \Psi_T^{-1} \sum_{i=1}^{N} \left[\theta_i \left(\mathbf{X}_i' \bar{\mathbf{M}}_w \boldsymbol{\xi}_i\right) - \gamma_i \theta_i \left(\mathbf{X}_i' \bar{\mathbf{M}}_w \bar{\boldsymbol{\xi}}_w\right)\right], \tag{A.80}$$

and to establish the consistency of $\hat{\mathbf{b}}$ it is sufficient to show that

$$\sum_{i=1}^{N} \left[w_i \left(\mathbf{X}_i' \bar{\mathbf{M}}_w \boldsymbol{\xi}_i\right) - \gamma_i w_i \left(\mathbf{X}_i' \bar{\mathbf{M}}_w \bar{\boldsymbol{\xi}}_w\right)\right] \xrightarrow{p} 0, \text{ for a fixed } T \text{ as } N \to \infty. \tag{A.81}$$

Consider first the second term in this sum[23]

$$\begin{aligned} E\left\|\sum_{i=1}^{N} \gamma_i \theta_i \left(\frac{\mathbf{X}_i' \bar{\mathbf{M}}_w \bar{\boldsymbol{\xi}}_w}{T}\right)\right\| &\leq \sum_{i=1}^{N} |\gamma_i|\,|\theta_i|\, E\left\|\frac{\mathbf{X}_i' \bar{\mathbf{M}}_w \bar{\boldsymbol{\xi}}_w}{T}\right\| \\ &\leq \sum_{i=1}^{N} |\gamma_i|\,|\theta_i|\, E\left(\frac{\sum_{t=1}^{T} \mathbf{x}_{it}' \mathbf{x}_{it}}{T}\right)^{1/2} E\left(\frac{\sum_{t=1}^{T} \bar{\xi}_{tw}^2}{T}\right)^{1/2} \\ &\leq \sum_{i=1}^{N} |\gamma_i|\,|\theta_i| \left[E\left(\frac{\sum_{t=1}^{T} \mathbf{x}_{it}' \mathbf{x}_{it}}{T}\right)\right]^{1/2} \left[E\left(\frac{\sum_{t=1}^{T} \bar{\xi}_{tw}^2}{T}\right)\right]^{1/2} \end{aligned}$$

Under assumptions 1(a), 2(a), and 5(a), for a fixed T, $E\left(\frac{\sum_{t=1}^{T} \mathbf{x}_{it}' \mathbf{x}_{it}}{T}\right) \leq K_1$, $|\gamma_i| < K_2$, and $\sum_{i=1}^{N} |\theta_i| \leq K_3$, for some positive constants K_i, $i = 1, 2, 3$,respectively and

$$E\left\|\sum_{i=1}^{N} \gamma_i \theta_i \left(\frac{\mathbf{X}_i' \bar{\mathbf{M}}_w \bar{\boldsymbol{\xi}}_w}{T}\right)\right\| \leq K_1 K_2 K_3 \ \left[E\left(\frac{\sum_{t=1}^{T} \bar{\xi}_{tw}^2}{T}\right)\right]^{1/2}. \tag{A.82}$$

[23]Note that $\left\|\mathbf{X}_i' \bar{\mathbf{M}}_w \bar{\boldsymbol{\xi}}_w\right\| \leq \left\|\mathbf{X}_i' \bar{\mathbf{M}}_w\right\| \left\|\bar{\boldsymbol{\xi}}_w\right\| \leq \left\|\mathbf{X}_i\right\| \left\|\bar{\boldsymbol{\xi}}_w\right\|$, and by assumption ξ_{it} and $\mathbf{x}_{it'}$ are independently distributed for all i, t and t'. Also for a positive random variable X with a finite second-order moment, by Jensen inequality we have $E(X^{1/2}) \leq [E(X)]^{1/2}$.

Therefore, using (A.10), we have

$$\lim_{N\rightarrow\infty} E\left\|\sum_{i=1}^{N}\gamma_i\theta_i\left(\mathbf{X}_i'\bar{\mathbf{M}}_w\bar{\boldsymbol{\xi}}_w\right)\right\| = 0, \text{ for a fixed } T,$$

and

$$\sum_{i=1}^{N}\gamma_i\theta_i\left(\mathbf{X}_i'\bar{\mathbf{M}}_w\bar{\boldsymbol{\xi}}_w\right) \xrightarrow{p} 0, \text{ as } N\rightarrow\infty. \tag{A.83}$$

Consider now the first term in (A.81) and using (A.73) note that

$$\sum_{i=1}^{N}\theta_i\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\boldsymbol{\xi}_i}{T} = \sum_{i=1}^{N}\theta_i\frac{\mathbf{X}_i'\Delta_N\boldsymbol{\xi}_i}{T} + \sum_{i=1}^{N}\theta_i\frac{\mathbf{X}_i'\mathbf{M}_\mu\boldsymbol{\xi}_i}{T}, \tag{A.84}$$

and (see the derivation of (**??**))

$$\left\|\sum_{i=1}^{N}\theta_i\frac{\mathbf{X}_i'\Delta_N\boldsymbol{\xi}_i}{T}\right\| \leq \|\Delta_N\|\sum_{i=1}^{N}|\gamma_i|\,|\theta_i|\left(\frac{\sum_{t=1}^{T}\mathbf{x}_{it}'\mathbf{x}_{it}}{T}\right)^{1/2}\left(\frac{\sum_{t=1}^{T}\xi_{it}^2}{T}\right)^{1/2}. \tag{A.85}$$

Under assumptions 1(a), 2(a), and 5(a) and using (A.77) it now follows that

$$\left\|\sum_{i=1}^{N}\theta_i\frac{\mathbf{X}_i'\Delta_N\boldsymbol{\xi}_i}{T}\right\| \xrightarrow{p} 0, \text{ as } N\rightarrow\infty, \text{ for all } T, \tag{A.86}$$

and consistency of $\hat{\mathbf{b}}$ is established provided

$$\left\|\sum_{i=1}^{N}\theta_i\frac{\mathbf{X}_i'\mathbf{M}_\mu\boldsymbol{\xi}_i}{T}\right\| \xrightarrow{p} 0. \tag{A.87}$$

However, since by assumption $\xi_{it'}$ is distributed independently of $(\mathbf{x}_{it}', \mathbf{g}_{t\mu}')$ for all i , t, and t', we have

$$Var\left(\sum_{i=1}^{N}\theta_i\frac{\mathbf{X}_i'\mathbf{M}_\mu\boldsymbol{\xi}_i}{T}\right) = \sum_{i=1}^{N}\theta_i^2\left\{\frac{\sigma_i^2}{T}E\left(\frac{\mathbf{X}_i'\mathbf{M}_\mu\mathbf{X}_i}{T}\right) + E\left[\left(\frac{\mathbf{X}_i'\mathbf{M}_\mu\mathbf{X}_i}{T}\right)\Omega\left(\frac{\mathbf{X}_i'\mathbf{M}_\mu\mathbf{X}_i}{T}\right)\right]\right\}. \tag{A.88}$$

But $E\left(\mathbf{X}_i'\mathbf{M}_\mu\mathbf{X}_i\right) \leq E\left(\mathbf{X}_i'\mathbf{X}_i\right)$, and under assumptions 1(a), 3, 4 and 5(a) it is easily verified that[24]

$$Var\left(\sum_{i=1}^{N}\theta_i\frac{\mathbf{X}_i'\mathbf{M}_\mu\boldsymbol{\xi}_i}{T}\right) \leq K\left(\sum_{i=1}^{N}\theta_i^2\right) \rightarrow 0, \text{ as } N\rightarrow\infty \text{ for all } T.$$

Hence $\sum_{i=1}^{N}\theta_i\mathbf{X}_i'\mathbf{M}_\mu\boldsymbol{\xi}_i \xrightarrow{p} \lim_{N\rightarrow\infty}\sum_{i=1}^{N}\theta_i E\left(\mathbf{X}_i'\mathbf{M}_\mu\boldsymbol{\xi}_i\right) = \mathbf{0}$, as required.[25]

[24]Note that using (2.14) we have $\sum_{i=1}^{N}\theta_i^2 = O(1/N)$.

[25]Recall that ξ_{it} are distributed with zero means independntly of $\mathbf{x}_{it'}$, $\mathbf{z}_{t'}$ and $\mathbf{f}_{t'}$, for all i, t, and t'.

To derive the asymptotic distribution of $\hat{\mathbf{b}}$, using (4.40) and (A.79), we first note that

$$\left(\sum_{i=1}^{N}\theta_i^2\right)^{-1/2}\left(\hat{\mathbf{b}}-\boldsymbol{\beta}\right) \overset{d}{\sim} \Psi_T^{-1}\sum_{i=1}^{N}\left[\tilde{\theta}_i\mathbf{X}_i'\bar{\mathbf{M}}_w\boldsymbol{\xi}_i-\gamma_i\tilde{\theta}_i\left(\mathbf{X}_i'\bar{\mathbf{M}}_w\bar{\boldsymbol{\xi}}_w\right)\right], \tag{A.89}$$

where

$$\tilde{\theta}_i=\frac{\theta_i}{\sqrt{\sum_{j=1}^{N}\theta_j^2}}=O\left(\frac{1}{\sqrt{N}}\right). \tag{A.90}$$

But under assumption 2(b)

$$\begin{aligned}\sum_{i=1}^{N}\gamma_i\tilde{\theta}_i\left(\mathbf{X}_i'\bar{\mathbf{M}}_w\bar{\boldsymbol{\xi}}_w\right) &= \gamma\sum_{i=1}^{N}\tilde{\theta}_i\left(\mathbf{X}_i'\bar{\mathbf{M}}_w\bar{\boldsymbol{\xi}}_w\right)+\sum_{i=1}^{N}\eta_i\tilde{\theta}_i\left(\mathbf{X}_i'\bar{\mathbf{M}}_w\bar{\boldsymbol{\xi}}_w\right)\\ &= \gamma\left(\tilde{\mathbf{X}}_\theta'\bar{\mathbf{M}}_w\bar{\boldsymbol{\xi}}_w\right)+\sum_{i=1}^{N}\eta_i\tilde{w}_i\left(\mathbf{X}_i'\bar{\mathbf{M}}_w\bar{\boldsymbol{\xi}}_w\right),\end{aligned} \tag{A.91}$$

where $\tilde{\mathbf{X}}_\theta=\sum_{i=1}^{N}\tilde{\theta}_i\mathbf{X}_i$. However, it is easily seen that in general $\gamma\left(\tilde{\mathbf{X}}_\theta'\bar{\mathbf{M}}_w\bar{\boldsymbol{\xi}}_w\right)$ does not vanish even as $N\to\infty$. Note that, by assumption $\gamma\neq 0$, and under (2.17) we have

$$\tilde{\mathbf{X}}_\theta'\bar{\mathbf{M}}_w\bar{\boldsymbol{\xi}}_w=\left(\sum_{i=1}^{N}\tilde{\theta}_i\Phi_{i3}'\right)\boldsymbol{\chi}'\bar{\mathbf{M}}_w\bar{\boldsymbol{\xi}}_w+\left(\sum_{i=1}^{N}\tilde{\theta}_i\mathbf{V}_i'\right)\bar{\mathbf{M}}_w\bar{\boldsymbol{\xi}}_w, \tag{A.92}$$

where $\boldsymbol{\chi}'=(\chi_1,\chi_2,...,\chi_N)$, and $\mathbf{V}_i'=(\mathbf{v}_{i1},\mathbf{v}_{i2},...,\mathbf{v}_{iT})$. Whilst it is possible to show that

$$\left(\sum_{i=1}^{N}\tilde{\theta}_i\mathbf{V}_i'\right)\bar{\mathbf{M}}_w\bar{\boldsymbol{\xi}}_w\overset{p}{\to}\mathbf{0},$$

the same can not be said about the first term of (A.92). This is because $\sum_{i=1}^{N}\tilde{\theta}_i\Phi_{i3}=O(\sqrt{N})$ and for this term to tend to zero as $N\to\infty$, we must have $\sqrt{N}\left(\boldsymbol{\chi}'\bar{\mathbf{M}}_w\bar{\boldsymbol{\xi}}_w\right)\overset{p}{\to}\mathbf{0}$, which in general does not seem possible. However, the problem can be avoided if we set $\theta_i=w_i$. Under this restriction

$$\tilde{\mathbf{X}}_\theta=\frac{\sum_{i=1}^{N}w_i\mathbf{X}_i}{\sqrt{\sum_{j=1}^{N}w_j^2}}=\frac{\bar{\mathbf{X}}_w}{\sqrt{\sum_{j=1}^{N}w_j^2}},$$

and we have (identically)

$$\tilde{\mathbf{X}}_\theta'\bar{\mathbf{M}}_w\bar{\boldsymbol{\xi}}_w=\frac{\bar{\mathbf{X}}_w'\bar{\mathbf{M}}_w\bar{\boldsymbol{\xi}}_w}{\sqrt{\sum_{j=1}^{N}w_j^2}}=\mathbf{0}.$$

Therefore, under $\theta_i=w_i$ we have

$$\left(\sum_{i=1}^{N}w_i^2\right)^{-1/2}\left(\hat{\mathbf{b}}-\boldsymbol{\beta}\right)\overset{d}{\sim}\Psi_T^{-1}\left[\sum_{i=1}^{N}\tilde{w}_i\left(\mathbf{X}_i'\bar{\mathbf{M}}_w\boldsymbol{\xi}_i\right)-\sum_{i=1}^{N}\eta_i\tilde{w}_i\left(\mathbf{X}_i'\bar{\mathbf{M}}_w\bar{\boldsymbol{\xi}}_w\right)\right]. \tag{A.93}$$

Consider now

$$\sum_{i=1}^{N} \eta_i \tilde{w}_i \left(\frac{\mathbf{X}_i' \bar{\mathbf{M}}_w \bar{\boldsymbol{\xi}}_w}{T} \right) = \left(\frac{\sum_{i=1}^{N} \eta_i \tilde{w}_i \mathbf{X}_i'}{\sqrt{T}} \right) \left(\frac{\bar{\mathbf{M}}_w \bar{\boldsymbol{\xi}}_w}{\sqrt{T}} \right), \tag{A.94}$$

and note that under assumption 2(b), $E\left(\sum_{i=1}^{N} \eta_i \tilde{w}_i \mathbf{X}_i'\right) = \mathbf{0}$, and hence

$$Var\left[\sum_{i=1}^{N} \frac{\eta_i \tilde{w}_i \mathbf{X}_i'}{\sqrt{T}}\right] = \sigma_\eta^2 \sum_{i=1}^{N} \tilde{w}_i^2 E\left(\frac{\mathbf{X}_i' \mathbf{X}_i}{T}\right), \tag{A.95}$$

which is bounded in N (and T), namely $T^{-1/2} \sum_{i=1}^{N} \eta_i \tilde{w}_i \mathbf{X}_i' = O_p(1)$. Also

$$E\left\| \frac{\bar{\mathbf{M}}_w \bar{\boldsymbol{\xi}}_w}{\sqrt{T}} \right\| \leq K \left[E\left(\frac{\sum_{t=1}^{T} \bar{\xi}_{it}^2}{T} \right) \right]^{1/2},$$

and by (A.10) we have

$$E\left(\frac{\sum_{t=1}^{T} \bar{\xi}_{it}^2}{T} \right) \to 0.$$

(for all T) and as $N \to \infty$. Thus for all T, and as $N \to \infty$

$$\sum_{i=1}^{N} \eta_i \tilde{w}_i \left(\frac{\mathbf{X}_i' \bar{\mathbf{M}}_w \bar{\boldsymbol{\xi}}_w}{T} \right) \xrightarrow{p} \mathbf{0}. \tag{A.96}$$

Similarly,

$$\sum_{i=1}^{N} \tilde{w}_i \left(\frac{\mathbf{X}_i' \bar{\mathbf{M}}_w \boldsymbol{\xi}_i}{T} \right) = \sum_{i=1}^{N} \tilde{w}_i \left(\frac{\mathbf{X}_i' \boldsymbol{\xi}_i}{T} \right) - \sum_{i=1}^{N} \tilde{w}_i \left(\frac{\mathbf{X}_i' \bar{\mathbf{H}}_w}{T} \right) \left(\frac{\bar{\mathbf{H}}_w \bar{\mathbf{H}}_w'}{T} \right)^{-1} \left(\frac{\bar{\mathbf{H}}_w' \boldsymbol{\xi}_i}{T} \right),$$

and using (A.24), (A.26) and (A.27) we have (for all T and as $N \to \infty$)

$$\sum_{i=1}^{N} \tilde{w}_i \left(\frac{\mathbf{X}_i' \bar{\mathbf{M}}_w \boldsymbol{\xi}_i}{T} \right) - \sum_{i=1}^{N} \tilde{w}_i \left(\frac{\mathbf{X}_i' \mathbf{M}_\mu \boldsymbol{\xi}_i}{T} \right) \xrightarrow{p} \mathbf{0}. \tag{A.97}$$

Using (A.96) and (A.97) in (A.93) we now have (for a fixed T)

$$\left(\sum_{i=1}^{N} w_i^2 \right)^{-1/2} \left(\hat{\mathbf{b}} - \boldsymbol{\beta} \right) \overset{d}{\sim} \Psi_T^{-1} \sum_{i=1}^{N} \tilde{w}_i \mathbf{X}_i' \mathbf{M}_\mu \boldsymbol{\xi}_i. \tag{A.98}$$

Also since by assumption $\xi_{it'}$ is a mean zero stationary process distributed independently across i, and of $\mathbf{x}_{it}$ and $\mathbf{g}_{t\mu}$ for all i, t and t', then by application of standard central limit theorem

$$\sum_{i=1}^{N} \tilde{w}_i \mathbf{X}_i' \mathbf{M}_\mu \boldsymbol{\xi}_i \xrightarrow{d} N(\mathbf{0}, \mathbf{R}_T), \text{ for a fixed } T \text{ as } N \to \infty, \tag{A.99}$$

where

$$\mathbf{R}_T = \text{plim}_{N\to\infty}\left\{\sum_{i=1}^{N}\tilde{w}_i^2\mathbf{X}_i'\mathbf{M}_\mu Var(\boldsymbol{\xi}_i)\mathbf{M}_\mu\mathbf{X}_i\right\}, \tag{A.100}$$

and

$$Var(\boldsymbol{\xi}_i) = \sigma_i^2\mathbf{I}_T + \mathbf{X}_i\Omega\mathbf{X}_i'. \tag{A.101}$$

Therefore

$$\left(\sum_{i=1}^{N}w_i^2\right)^{-1/2}\left(\hat{\mathbf{b}}-\boldsymbol{\beta}\right)\xrightarrow{d}N(\mathbf{0},\Psi_T^{-1}\mathbf{R}_T\Psi_T^{-1}),\text{ for a fixed }T\text{ as }N\to\infty. \tag{A.102}$$

Also using (A.101) in (A.100) we have,

$$\mathbf{R}_T = \text{plim}_{N\to\infty}\left\{\sum_{i=1}^{N}\tilde{w}_i^2\left[\sigma_i^2\Psi_{iT}+\Psi_{iT}\Omega\Psi_{iT}\right]\right\},\text{ with }\Psi_{iT}=\mathbf{X}_i'\mathbf{M}_\mu\mathbf{X}_i.$$

It is also easily verified that under our assumptions and noting that $\tilde{w}_i^2 = O(1/N)$, $\mathbf{R}_T$ exists and is finite.

Part (b) - **Joint N and T asymptotic**. First note that

$$\sum_{i=1}^{N}\theta_i\left(\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{X}_i}{T}\right) = \sum_{i=1}^{N}\theta_i\left(\frac{\mathbf{X}_i'\Delta_N\mathbf{X}_i}{T}\right)+\sum_{i=1}^{N}\theta_i\left(\frac{\mathbf{X}_i'\mathbf{M}_\mu\mathbf{X}_i}{T}\right). \tag{A.103}$$

where Δ_N is defined by (A.73). Also using (A.74) and (A.75) we have

$$\left\|\sum_{i=1}^{N}\theta_i\frac{\mathbf{X}_i'\Delta_N\mathbf{X}_i}{T}\right\| \leq \|\Delta_N\|\sum_{i=1}^{N}|\theta_i|\left(\frac{\sum_{t=1}^{T}\mathbf{x}_{it}'\mathbf{x}_{it}}{T}\right).$$

By assumption 1(a) and condition 5(a)

$$\text{plim}_{(N,T)\xrightarrow{j}\infty}\sum_{i=1}^{N}|\theta_i|\left(\frac{\sum_{t=1}^{T}\mathbf{x}_{it}'\mathbf{x}_{it}}{T}\right) < K,$$

and by (A.77) $\|\Delta_N\|\xrightarrow{p}0$ as $(N,T)\xrightarrow{j}\infty$. Hence

$$\text{plim}_{(N,T)\xrightarrow{j}\infty}\left\|\sum_{i=1}^{N}\theta_i\frac{\mathbf{X}_i'\Delta_N\mathbf{X}_i}{T}\right\| = 0. \tag{A.104}$$

Using this result in (A.103) now yields

$$\begin{aligned}\text{plim}_{(N,T)\xrightarrow{j}\infty}\left[\sum_{i=1}^{N}\theta_i\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\mathbf{X}_i}{T}\right] &= \text{plim}_{(N,T)\xrightarrow{j}\infty}\left[\sum_{i=1}^{N}\theta_i\frac{\mathbf{X}_i'\mathbf{M}_\mu\mathbf{X}_i}{T}\right]\\ &= \lim_{N\to\infty}\left(\sum_{i=1}^{N}\theta_i\Psi_i\right) = \Psi,\end{aligned}$$

where Ψ_i is a positive definite matrix defined by (A.53). Note that Ψ is also a positive definite matrix (by assumption 5(c)). Hence from (4.40) we have

$$\text{plim}_{(N,T)\overset{j}{\to}\infty}(\hat{\mathbf{b}}) - \boldsymbol{\beta} = \Psi^{-1}\left\{\text{plim}_{(N,T)\overset{j}{\to}\infty}\sum_{i=1}^{N}\left[\theta_i\left(\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\boldsymbol{\xi}_i}{T}\right) - \gamma_i\theta_i\left(\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\bar{\boldsymbol{\xi}}_w}{T}\right)\right]\right\}.$$

Using (A.82) and (A.10) we have

$$\lim_{(N,T)\overset{j}{\to}\infty} E\left\|\sum_{i=1}^{N}\theta_i w_i\left(\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\bar{\boldsymbol{\xi}}_w}{T}\right)\right\| = \mathbf{0}. \tag{A.105}$$

Similarly, using (A.84), (A.86) and (A.88)

$$\lim_{(N,T)\overset{j}{\to}\infty} Var\left(\sum_{i=1}^{N}\theta_i\frac{\mathbf{X}_i'\mathbf{M}_\mu\boldsymbol{\xi}_i}{T}\right) = \mathbf{0}, \tag{A.106}$$

and

$$plim_{(N,T)\overset{j}{\to}\infty}\sum_{i=1}^{N}\theta_i\left(\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\boldsymbol{\xi}_i}{T}\right) = \mathbf{0}. \tag{A.107}$$

Hence, $\text{plim}_{(N,T)\overset{j}{\to}\infty}(\hat{\mathbf{b}}) = \boldsymbol{\beta}$, as required.

Derivation of the asymptotic distribution of $\hat{\mathbf{b}}$ under $(N,T)\overset{j}{\to}\infty$ can be carried out along similar lines as in the previous case, except that the nature of the distribution and its rate of convergence depends on whether $\boldsymbol{\beta}_i$'s are heterogeneous, namely whether Ω is a positive definite matrix or $\Omega = \mathbf{0}$.[26] However, $\hat{\mathbf{b}}$ continues to be consistent for $\boldsymbol{\beta}$ irrespective of whether $\Omega = \mathbf{0}$ or not.

First recall that under $\theta_i = w_i$ and assumption 2(b) we have

$$\sum_{i=1}^{N}\gamma_i\tilde{w}_i\left(\mathbf{X}_i'\bar{\mathbf{M}}_w\bar{\boldsymbol{\xi}}_w\right) = \sum_{i=1}^{N}\eta_i\tilde{w}_i\left(\mathbf{X}_i'\bar{\mathbf{M}}_w\bar{\boldsymbol{\xi}}_w\right). \tag{A.108}$$

Therefore, using (4.40) and (A.105),

$$\left(\sum_{i=1}^{N}w_i^2\right)^{-1/2}\left(\hat{\mathbf{b}} - \boldsymbol{\beta}\right) \overset{d}{\sim} \Psi^{-1}\left[\sum_{i=1}^{N}\tilde{w}_i\left(\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\boldsymbol{\xi}_i}{T}\right) - \sum_{i=1}^{N}\eta_i\tilde{w}_i\left(\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\bar{\boldsymbol{\xi}}_w}{T}\right)\right]. \tag{A.109}$$

Also using (A.96) and (A.97) we have

$$\left(\sum_{i=1}^{N}w_i^2\right)^{-1/2}\left(\hat{\mathbf{b}} - \boldsymbol{\beta}\right) \overset{d}{\sim} \Psi^{-1}\sum_{i=1}^{N}\tilde{w}_i\left(\frac{\mathbf{X}_i'\mathbf{M}_\mu\boldsymbol{\xi}_i}{T}\right), \text{ as } (N,T)\overset{j}{\to}\infty. \tag{A.110}$$

[26]Here we do not consider intermediate cases where a sub-set of $\boldsymbol{\beta}_i$ could be homogenous.

As with derivation of (A.99), in the case where Ω is a positive definite matrix it is easily seen that with $(N,T) \xrightarrow{j} \infty$,

$$\sum_{i=1}^{N} \tilde{w}_i \left(\frac{\mathbf{X}_i' \mathbf{M}_\mu \boldsymbol{\xi}_i}{T} \right) \xrightarrow{d} N(\mathbf{0}, \mathbf{R}), \text{ as } (N,T) \xrightarrow{j} \infty, \tag{A.111}$$

$$\begin{aligned} \mathbf{R} &= \text{plim}_{(N,T) \xrightarrow{j} \infty} \left\{ \sum_{i=1}^{N} \tilde{w}_i^2 \left[\sigma_i^2 \left(\frac{\mathbf{X}_i' \mathbf{M}_\mu \mathbf{X}_i}{T^2} \right) + \left(\frac{\mathbf{X}_i' \mathbf{M}_\mu \mathbf{X}_i}{T} \right) \Omega \left(\frac{\mathbf{X}_i' \mathbf{M}_\mu \mathbf{X}_i}{T} \right) \right] \right\}, \\ &= \lim_{N \to \infty} \left\{ \sum_{i=1}^{N} \tilde{w}_i^2 \Psi_i \Omega \Psi_i \right\}, \end{aligned}$$

where Ω and Ψ_i are non-stochastic positive definite matrices, and $\tilde{w}_i^2 = O\left(N^{-1}\right)$. Hence, $\mathbf{R}$ exists and is also a positive definite matrix. Using (A.111) in (A.110) we have the desired result, namely

$$\left(\sum_{i=1}^{N} w_i^2 \right)^{-1/2} \left(\hat{\mathbf{b}} - \boldsymbol{\beta} \right) \xrightarrow{d} N(\mathbf{0}, \Psi^{-1} \mathbf{R} \Psi^{-1}). \tag{A.112}$$

Consider now the homogeneous slope case where $\Omega = \mathbf{0}$. In this case $\hat{\mathbf{b}}$ converges to $\boldsymbol{\beta}$ at a faster rate, also helped by T. To see this note that in this case $\boldsymbol{\xi}_i = \boldsymbol{\varepsilon}_i$, and $Var\left(\boldsymbol{\varepsilon}_i\right) = \sigma_i^2 \mathbf{I}_T$. Using (A.109) we have

$$\left(\frac{\sum_{i=1}^{N} w_i^2}{T} \right)^{-1/2} \left(\hat{\mathbf{b}} - \boldsymbol{\beta} \right) \overset{d}{\sim} \Psi^{-1} \left[\sum_{i=1}^{N} \tilde{w}_i \left(\frac{\mathbf{X}_i' \bar{\mathbf{M}}_w \boldsymbol{\varepsilon}_i}{\sqrt{T}} \right) - \sum_{i=1}^{N} \eta_i \tilde{w}_i \left(\frac{\mathbf{X}_i' \bar{\mathbf{M}}_w \bar{\boldsymbol{\varepsilon}}_w}{\sqrt{T}} \right) \right]. \tag{A.113}$$

As before first consider

$$\sum_{i=1}^{N} \eta_i \tilde{w}_i \left(\frac{\mathbf{X}_i' \bar{\mathbf{M}}_w \bar{\boldsymbol{\varepsilon}}_w}{\sqrt{T}} \right) = \frac{\sum_{i=1}^{N} \eta_i \tilde{w}_i \mathbf{X}_i' \bar{\boldsymbol{\varepsilon}}_w}{\sqrt{T}} - \left(\frac{\sum_{i=1}^{N} \eta_i \tilde{w}_i \mathbf{X}_i' \bar{\mathbf{H}}_w}{T} \right) \left(\frac{\bar{\mathbf{H}}_w' \bar{\mathbf{H}}_w}{T} \right)^{-1} \left(\frac{\bar{\mathbf{H}}_w' \bar{\boldsymbol{\varepsilon}}_w}{\sqrt{T}} \right), \tag{A.114}$$

note that $E\left(\sum_{i=1}^{N} \eta_i \tilde{w}_i \mathbf{X}_i' \bar{\boldsymbol{\varepsilon}}_w \right) = 0$, and

$$Var\left(\frac{\sum_{i=1}^{N} \eta_i \tilde{w}_i \mathbf{X}_i' \bar{\boldsymbol{\varepsilon}}_w}{\sqrt{T}} \right) = \sigma_\eta^2 \left(\sum_{i=1}^{N} \sigma_i^2 w_i^2 \right) \left[\sum_{i=1}^{N} \tilde{w}_i^2 E\left(\frac{\mathbf{X}_i' \mathbf{X}_i}{T} \right) \right].$$

Since $\tilde{w}_i^2 = O\left(N^{-1}\right)$, $E\left(\frac{\mathbf{X}_i' \mathbf{X}_i}{T}\right)$ is bounded in T for each i, we have $\sum_{i=1}^{N} \tilde{w}_i^2 E\left(\frac{\mathbf{X}_i' \mathbf{X}_i}{T}\right) = O(1)$ for all N and T, and noting that $\sum_{i=1}^{N} w_i^2 = O\left(N^{-1}\right)$ then as $(N,T) \xrightarrow{j} \infty$

$$Var\left(\frac{\sum_{i=1}^{N} \eta_i \tilde{w}_i \mathbf{X}_i' \bar{\boldsymbol{\varepsilon}}_w}{\sqrt{T}} \right) \to \mathbf{0},$$

and we have

$$\frac{\sum_{i=1}^{N} \eta_i \tilde{w}_i \mathbf{X}_i' \bar{\boldsymbol{\varepsilon}}_w}{\sqrt{T}} \xrightarrow{p} \mathbf{0}, \text{ as } (N,T) \xrightarrow{j} \infty.$$

Also using (A.26) and (A.52) we have

$$\frac{\mathbf{X}_i'\bar{\mathbf{H}}_w}{T} = \Sigma_{\mu x_i}'\mathbf{A}_w' + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right),$$

and under assumptions 1(a) and 2(b), it is easily verified that

$$T^{-1}\left(\sum_{i=1}^{N}\eta_i\tilde{w}_i\mathbf{X}_i'\bar{\mathbf{H}}_w\right) = O_p(1) \text{ for all } N \text{ and } T.$$

Finally, using (A.24) and (A.27)

$$\left(\frac{\bar{\mathbf{H}}_w'\bar{\mathbf{H}}_w}{T}\right)^{-1}\left(\frac{\bar{\mathbf{H}}_w'\bar{\boldsymbol{\varepsilon}}_w}{\sqrt{T}}\right) = O_p\left(\frac{\sqrt{T}}{N}\right) + O_p\left(\frac{1}{\sqrt{N}}\right),$$

and we have

$$\sum_{i=1}^{N}\eta_i\tilde{w}_i\left(\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\bar{\boldsymbol{\varepsilon}}_w}{\sqrt{T}}\right)\frac{\sum_{i=1}^{N}\eta_i\tilde{w}_i\mathbf{X}_i'\bar{\boldsymbol{\varepsilon}}_w}{\sqrt{T}} \xrightarrow{p} \mathbf{0},$$

provided that $\sqrt{T}/N \to 0$, as $(N,T) \xrightarrow{j} \infty$. Under these conditions

$$\left(\frac{\sum_{i=1}^{N}w_i^2}{T}\right)^{-1/2}\left(\hat{\mathbf{b}}-\boldsymbol{\beta}\right) \overset{d}{\sim} \Psi^{-1}\sum_{i=1}^{N}\tilde{w}_i\left(\frac{\mathbf{X}_i'\bar{\mathbf{M}}_w\boldsymbol{\varepsilon}_i}{\sqrt{T}}\right),$$

and using similar line of reasoning as above we have

$$\left(\frac{\sum_{i=1}^{N}w_i^2}{T}\right)^{-1/2}\left(\hat{\mathbf{b}}-\boldsymbol{\beta}\right) \xrightarrow{d} N(\mathbf{0},\Psi^{-1}\dot{\mathbf{R}}\Psi^{-1}),$$

where

$$\dot{\mathbf{R}} = \lim_{N\to\infty}\left(\sum_{i=1}^{N}\tilde{w}_i^2\sigma_i^2\Psi_i\right).$$

References

[1] Ahn, S.G., Y,H, Lee and Schmidt, P. (2001), GMM Estimation of Linear Panel Data Models with Time-varying Individual Effects, *Journal of Econometrics*, 102, 219-255.

[2] Anselin, L. (2001), "Spatial Econometrics", in B. Baltagi (ed.), *A Companion to Theoretical Econometrics*, Blackwell, Oxford.

[3] Bai, J. and Ng, S. (2002), "Determining the Number of Factors in Approximate Factor Models", *Econometrica*, 70, 191-221.

[4] Binder, M., C. Hsiao, Pesaran, M.H. (2002), "Estimation and Inference in Short Panel Vector Autoregressions with Unit Roots and Cointegration", University of Cambridge DAE Working Paper No.0003.

[5] Coakley, J., Fuerts, A., and Smith, R.P. (2002), "A Principal Components Approach to Cross-Section Dependence in Panels", Unpublished manuscript, Birckbeck College, University of London.

[6] Conley, T.G. and Dupor, B. (2001), "A Spatial Analysis of Sectoral Complementarity", forthcoming in *Journal of Political Economy.*

[7] Conley, T.G. and Topa, G. (2002), "Socio-economic Distance and Spatial Patterns in Unemployment", forthcoming in *Journal of Applied Econometrics.*

[8] Forni, M., and Lippi, M. (1997), *Aggregation and the Microfoundations of Dynamic Macroeconomics*, Calrendon Press, Oxford.

[9] Forni, M., and Reichlin, L. (1998), "Let's Get Real: A Factor Analytical Approach to Disaggregated Business Cycle Dynamics", *Review of Economic Studies*, 65, 453-73.

[10] Hamilton, J.D. (1994), The Time Series Analysis, Princeton University Press, New Jersey.

[11] Holtz_Eakin, D, Newey, W.K., and Rosen, H., (1988), "Estimating Vector Autoregressions with Panel Data", *Econometrica*, 56, 1371-1395.

[12] Hsiao, C., Pesaran, M.H., and Tahmiscioglu, A.K. (1999), Bayes Estimation of Short-run coefficients in Dynamic Panel Data Models in C. Hsiao, K. Lahiri, L-F Lee and M.H. Pesaran (eds), *Analysis of Panels and Limited Dependent Variables: A Volume in Honour of G S Maddala*, Cambridge University Press, Cambridge.

[13] Kiefer, N.M. (1980), "A Time Series-Cross Section Model with Fixed Effects with an Intertemporal Factor Structure. Unpublished manuscript, Department of Economics, Cornell University.

[14] Lee, Y.H. (1991), "Panel Data Models with Multiplicative Individual and Time Effects_ Application to Compensation and Frontier Production Functions", Unpublished Ph.D. Dissertation, Mishigan State University.

[15] Lee, K.C., and Pesaran, M.H. (1993), "The Role of Sectoral Interactions in Wage Determination in the UK Economy", *The Economic Journal*, 103, 21-55.

[16] Pesaran, M.H. (2002), "Large Dynamic Heterogenous Panel Data Models with Cross-Section Dependence", under preparation.

[17] Pesaran, M.H. and Smith R.P. (1995) "Estimating Long-Run Relationships from Dynamic Heterogeneous Panels", *Journal of Econometrics*, 1995, 68, 79-113.

[18] Pesaran, M.H., Schuermann, T., and Weiner, S.M. (2001), "Modeling Regional Interdependencies using a Global Error-Correcting Macroeconomic Model", Mimeo, University of Cambridge.

[19] Phillips, P.C.B., and Sul, D. (2002), "Dynamic Panel Estimation and Homogeneity Testing Under Cross Section Dependence", Cowles Foundation Discussion Paper No. 1362, Yale University.

[20] Robertson, D. and Symons, J. (2000), "Factor Residuals in SUR Regressions: Estimating Panels Allowing for Cross Sectional Correlation", Unpublished manuscript, Faculty of Economics and Politics, University of Cambridge.

[21] Stock and Watson, M.W. (1998), "Diffusion Indexes", NBER Working Paper 6702.

[22] Swamy, P.A.V.B. (1970), Efficient Inference in Random Coefficient Regression Model, Econometrica, 38, 311-23.

[23] Zaffaroni, P. (2001), Contemporaneous Aggregation of Linear Dynamic Models in Large Economies, Banca d'Italia, Unpublished manuscript.

[24] Zellner, A. (1962), "An Efficient Method for Estimating Seemingly Unrelated Regressions and Tests of Aggregation Bias", Journal of American Statistical Association, 58, 977-992.